1. Overview

Where do galaxies come from? How do black holes form and grow? And what is the link, if any, between the central black hole and its host galaxy? These are among the deepest, most fundamental questions in contemporary astrophysics, and sets the scene for this proposal.

Black holes are omnipresent in our Universe, and black holes that are millions to billions of times the mass of our Sun, are ubiquitously found at the centers of galaxies, including our own Milky Way. Initially consider physical oddities, we now strongly suspect that these central, "supermassive" black holes have a profound affect on the galaxies that they live in. This is not surprising since the potential energy associated with mass accretion onto a supermassive black hole is comparable to that generated via the nuclear fusion in the galaxy's stars. However, the interaction and the physical processes involved in how this energy escapes the inner most regions of the galaxy and then interacts with the gas, dust, stars and dark matter, is currently very poorly understood theoretically, with observational data giving little insight on how to make key progress.

The field is poised for a fundamental and rapid change. The first data are now in hand that show changes on human timescales in external galaxies, with these new field-defining studies including projects led by the P.I. Moreover, a fleet of telescopes, detectors and missions are about to come online over the next few years that will leap-frog the quality and quantity of data we have available today. Over the course of the next 5-6 years, surveys and missions including SDSS-V[1], LSST[2], DESI, [3], 4MOST [4], and ESA *Euclid* [5], will see first light. Even more imminent is the launch of the *James Webb Space Telescope*[6].

The current state-of-the-art data samples have $\approx 10^6$ objects with one spectral epoch, or only $\sim$a few objects with repeat spectra (e.g, MacLeod, Ross et al. 2016; Ross et al. 2018). We plan to collate datasets so that the $\approx 10^6$ sample have light-curves and repeat spectra and in doing so, will kickstart the new field of Extragalactic Time-Domain Astrophysics.

This proposal has two broad but well-posed goals. First, we aim to elucidate, for the first time, how the energy directly associated with a supermassive black holes impacts the universal galaxy population. This will open up and explore the Variable Extragalactic Universe, bringing to bear the slew of new larget format "synoptic" telescopes. Second, we will discover new extragalactic phenomena. Things will go "bump in the middle of the night"; we just don't know what they are yet.

We will achieve this by leveraging several of the new, large-scale surveys that are coming online in the next few years. The goal is to connect the physical mechanisms from sub-parsec to cosmological scales, and to investigate the physical processes that link luminous AGN activity and the formation and evolution of massive galaxies. These critical observations are made by exploiting the large imaging and spectroscopic datasets that we will have available from the SDSS-V, DESI, 4MOST, LSST and ESA Euclid. We ask for the personnel to accomplish these ambitious, but achievable goals, along with the 'buy-in' to the facilites we need access to. We also ask for the computing and travel support that will directly enable this research.

The scope and remit of an ERC Consolidator grant will allow us to combine these data products in a manner that will not only establish the new state-of-the-art in extragalactic time-domain science, it will establish and kickstart the new field of extragalactic time-domain science itself.

2. Scientific Background

Quasars are extremely luminous extragalactic objects powered by gas falling onto supermassive black holes. They are detected in optical surveys, both via their photometric and spectra properties. Accretion onto a

[1] `www.sdss.org/future/`
[2] `lsst.org`
[3] `desi.lbl.gov`
[4] `4most.eu`
[5] `sci.esa.int/euclid/`
[6] `jwst.stsci.edu`

supermassive black hole is one of two major energy sources available to a galaxy. However, we do not know how this energy escapes the central regions where it is generated, and how it interacts with the galaxy at large.

In the Table below, we summarise the outstanding issues, and our novel investigations.

Outstanding Issues in Extragalactic Astrophysics	
Key issue	Novel investigation
THE PHYSICS OF ACCRETION	
Investigating "hot" and "cold" mode accretion in the quasar population; determining the rates and timescales, and characterising the Changing Look Quasar (CLQ) population.	Identifying and characterizing all the CLQs in DESI and SDSS-V.
Probing the inner parsec of the quasar central engine	Rapid analysis and response on LSST quasar light curves.
OBSCURED ACCRETION AND GALAXY FORMATION	
Establish the relative importance of major mergers, minor mergers, cold streams and secular evolution have towards the growth of SMBHs across cosmic time.	Deep imaging data from LSST combined with searching for post-starburst signatures in DESI, SDSS-V, 4MOST spectra. NIRcam and MIRI imaging from JWST.
Establishing the bolometric output and origin of IR emission, and determine presence of extreme outflows in the $z \sim 2-3$ quasar population.	MIRI MRS spectroscopy with JWST.
Establishing the range of SED parameter space the quasars occupy by a multi-wavelength multi-epoch "truth table dataset"	Building "The Stripe 82 Rosetta Stone" (SpIES, SHELA, VICS82, S82X, HSC; repeat optical observations from SDSS, DES)
Find the physical conditions under which SMBH grew at the epoch when most of the accretion and star formation in the Universe occurred ($z \sim 1-4$)	Perform a complete census of AGN across $z \sim 0-7$, focussing on $z = 1-4$ using medium-deep multi-wavelength datasets
GALAXY-SCALE FEEDBACK	
Establishing the theoretical impact of extreme outflows in the $z \sim 2-3$ quasar population	Hydro simulation modelling.
Understand how the accretion disks around black holes launch winds and outflows and determine how much energy these carry. Quantify the amount of "Maintenance/Jet/Kinetic" mode and "Transition/Radiative/Wind" mode feedback.	Identifying and characterizing all the CLQs in DESI and SDSS-V.

During its first phase of operation (2000–05), the Sloan Digital Sky Survey (SDSS) obtained spectra of 1 million galaxies in the local Universe. This dataset has become the *de facto* standard for understanding the present day galaxy population, and setting a boundary condition for all theoretical comparisons. *The paradigm changing success of the SDSS was due to it having 1e6 objects. We desire the same sample size and revolutionary understanding of the quasar population as the SDSS had with the $z \sim 0.1$ galaxy population.* Our proposal takes astrophysics into the 2020s, going from single objects samples, to surveys and samples of millions of objects leveraging these very large scale next generation missions, telescopes and their subsequent datasets.

3. The Future is Now: Project in larger context

GLOBAL MOTIVATION: At its heart, there are two major motivations for our project. The first is to gain a deep understanding into the physical mechanisms related to central engine black holes; their accretion disk physics, their dynamics on both human and galactic timescales and the role they might play in forming, and regulating the galaxy population. These are among the most prescient astrophysical questions of our time, and in an area where major breakthroughs are imminent.

The importance of this branch of astrophysics is already well establish in Europe and is a priority for the next two decades. This is demonstrated by noting that one of the two primary mission goals for the Advanced Telescope for High-ENergy Astrophysics (ATHENA) is answering the question "How do black holes grow and shape the Universe?". ATHENA is ESA's second L-class flagship mission, due for launch in 2028.

The second motivation is the massive, untapped and raw discovery space that the new experiments will open up, and the highly likely outcome of discovering something "brand new". It is somewhat tricky to say specifically what to expect, but the fact that e.g. LSST will deliver a dataset *so spectacularly* different both in sky coverage and time-sampling coverage, means the Universe would have to be an exceptionally boring place to not have brand new astronomical objects and astrophysical phenomena be discovered.

Maximising Science Returns from European priorities: Contemporary astronomy is a multi-national endeavor with many leading facilites being international collaborations. Although a project, with similar but much less ambitious science goals and return could be envisaged at the national level, the full discovery and break-through nature being described herein only comes to the fore when the data from the various international collaborations are combined intelligently. Critically data from leading European Southern Observatory (ESO) and European Space Agency (ESA) facilites will play a pivotal role here.

Although many of the "building blocks" for our science are already available (e.g. open source codes, database infrastructures, the methodology of catalog creation and combination) no one has yet to combine the data in the way we envisage. Moreover, the new datasets we desire to deliver our paradigm changing science are only coming online over the next 5 or so years.

Timing: *The timing for this proposal could not be better or more imperative.* The first of the data "firehoses" turns on in late 2019, with the full datastream from our key sources fully online around 2022. As such, with two years to use existing datasets as testbeds, we have the time to ramp-up our efforts, while also being able to take advantage of the initial data releases of all these new projects.

Data Science and Observational Astrophysics: Data science is a new interdisciplinary field of scientific methods to extract knowledge or insights from data in various forms, either structured or unstructured. It employs techniques and theories drawn from many fields within the broad areas of mathematics, statistics, information science, and computer science, in particular from the subdomains of machine learning, classification, cluster analysis, data mining, databases, and visualization. *Modern day observational astrophysicists are in all but name data scientists, and as such, this proposal is inherently interdisciplinary.*

The P.I. has become a world-leader in the field of extragalactic quasar observational astrophysics. Moreover, the University of Edinburgh is now poised to be an astronomical data centre nexus, with access to the two largest datasets in our proposal; LSST and *Euclid*. The P.I. has built their career on this science case, and has already been a P.I. of a science group (as part of a collaboration) with prodigious scientific output (400 published, peer-reviewed papers and counting).

4. Feasibility, Projects and Methodology

P.I.'s Experience and Track Record: The P.I. has an established track record of managing science teams and groups, e.g.,

- The P.I. was the Chair of the SDSS-III:BOSS Quasar Science Working Group, managing a group of senior professors, other postdocs and graduate students. The scientific yield from the BOSS Quasar Survey was extremely high with over 400 journal publications having used the BOSS Quasar catalogs and datasets.

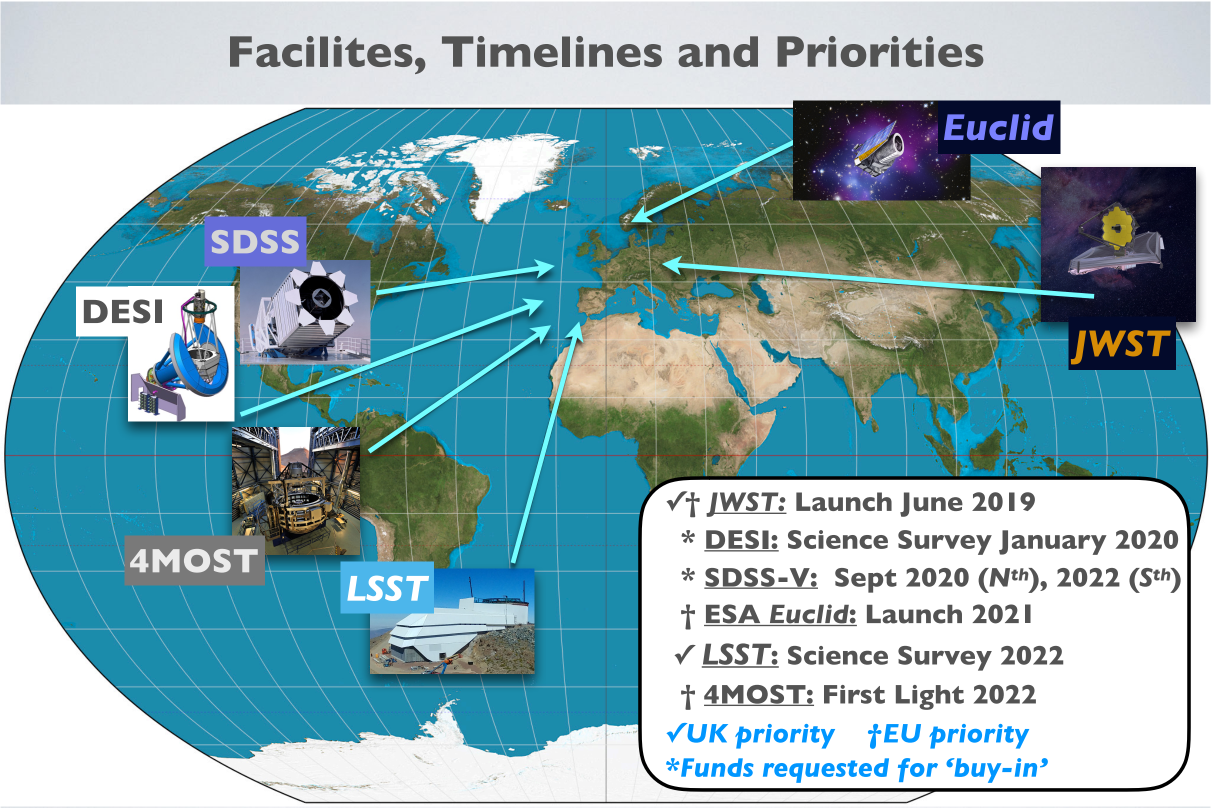

Figure 1:

- The P.I. spent a considerable amount of time in 2017 working as the Chief Data Scientist for a San Francisco Bay Area tech start-up[7].
- The P.I. is an STFC Ernest Rutherford Fellow, with a budget of €615,000 on award.

Breaking Down The Data Silos

to using advanced data analysis is not skill base or technology; it is simply access to the data. A data silo is a repository of fixed data that remains under the control of one department/collaboration and is isolated from the rest of the world, much like grain in a farm silo is closed off from outside elements. These silos are isolated islands of data, and they make it prohibitive to extract data and put it to other uses. They can arise for multiple reasons. In commercial enterprises, data remained siloed for monetary gain. However, in research environments, and *especially in contemporary observational astrophysics*, the data silos are open, but due to the lack of raw person-power, still remain uncombined. *The combination of P.I. and host institute means we are uniquely positioned to break down these astro-data silos for massively significant science gain.*

Algorithms and Methodology

Our algorithms and methodology is based on the latest machine-learning and data science techniques. This includes the "extreme deconvolution" 'XDQSO' technique[8]. `scikit-learn` is a Python module integrating classic machine learning algorithms in the scientific Python world (numpy, scipy, matplotlib). It aims to provide simple and efficient solutions to learning problems, accessible to everybody and reusable in various contexts. `github.com/astroML/sklearn_tutorial` and `github.com/jakevdp/PythonDataScienceHandbook` have full details.

[7]Due to legal immigration issues, this venture is no longer being pursued.
[8]`github.com/xdqso/xdqso`

Personel	Year 1	Year 2	Year 3	Year 4	Year 5
P.I.	90% time commitment				
PDRA 1	100% time commitment				
PDRA 2		100% time commitment			
PDRA 3		100% time commitment		100% time commitment	
PhD 1		100% time commitment			
Parental Leave	18 month equivalent				

One main ticket-item will be developing the *Event Broker Service* that will interface with the e.g. LSST databases, pipelines and transient stream. This is an already solved problem in many areas of DS and InfoSec, but we require a custom build in order to access the data streams necessary for our science goals.

Computing Requirements

With the availability of contemporary cloud based services (both commercial and non-for-profit), the infrastructure is essentially already in place for all but the most demanding of compute tasks. The rate limiting factor, in the vast majority of endevours, is (a) data access and (b) development.

The facilites available to me at an institute (e.g. IfA Cullen), university (e.g. "Edinburgh Compute and Data Facility" and at a national (The Hartree Centre) level will all be sufficient and utilized. The rate limiting factor will be how quickly and efficiently we can deploy our codes, and analysis, i.e. person-power.

Open Innovation, Open Science, Open to the World

The P.I. is an exceptionally strong, longtime and vocal supporter of "Open Access". All my codes, data[9], papers and proposals can be found at `github.com/d80b2t`. Indeed, this proposal itself is now at that location.

[9] where I am not breaking current data access agreements

References

[1]Kormendy & Ho, 2013, ARAA, 51, 511
[2]Kormendy, 2016, ASSL, 418, 431
[3]Alexander et al., 2012, NewAR, 56, 93
[4]King & Pounds, 2015, ARAA, 53, 115
[5]Heckman & Best, 2014, ARAA, 52, 589
[6]Naab & Ostriker, 2017, ARAA, 55, 59
[7]Netzer, 2015, ARAA, 53, 365
[8]Padovani, 2017, A&ARv, 25, 2
[9]Ata et al., 2017, arXiv1705.06373v2
[10]Solsar et al., 2013, JCAP, 04, 026
[11]Busca et al., 2013, A&A, 552, 96
[12]Delubac et al., 2015, A&A, 574, 59
[13]Bautista et al., 2017, A&A, 603, 12
[14]du Mas des Bourboux et al., 2017, arXiv1708.02225v3
[15]Font-Riber et al., 2014, JCAP, 05, 027
[16]Ross et al. 2015, MNRAS, 453, 3932
[17]Wright et al., 2010, AJ, 140, 1868
[18]Zakamska et al., 2016, MNRAS, 459, 3144
[19]Hamann et al., 2017, MNRAS, 464, 3431
[20]Timlin, Ross et al., 2016, ApJS, 225, 1
[21]Timlin, Ross et al., 2017, ApJ, *in prep.*
[22]LaMassa et al., 2015, ApJ, 800, 144
[23]Runnoe et al., 2016, MNRAS, 455, 1691
[24]Ruan et al, 2016, ApJ, 826, 188
[25]MacLeod, Ross et al., 2016, MNRAS, 457, 389
[26]Meisner et al., 2017, AJ, 153, 38
[27]Meisner et al., 2017, AJ, 154, 161
[28]Ross et al., 2017, Nat.As., *in prep.*
[29]Hopkins et al., 2006, ApJS, 163, 1
[30]Schlegel et al., 2011, arXiv:1106.1706v2
[31]Ross et al., 2009, ApJ, 697, 1634
[32]Lombriser & Taylor, 2016, JCAP, 03, 031
[33]Baker et al, 2017, arXiv1710.06394v1
[34]Watson et al., 2011, ApJ, 740, L49
[35]King et al., 2014, MNRAS, 441, 3454
[36]King et al., 2015, MNRAS, 453, 1701
[37]Shen et al., 2015, ApJS, 216, 4
[38]Hviding et al., 2017, arXiv1711.01269v1