Commercial Real Estate Data:
Towards Parity with Other Asset Classes

A Report on the progress of the Commercial Real Estate Data Alliance (CREDA)[1]

VERSION: January 5, 2016

[1] This report represents the joint efforts of current CREDA affiliates, listed in Appendix B of this document. We have greatly benefitted from multiple discussions with Richard Barkham, Brad Case, Deb Cloutier, Joe Emison, Lynn Fisher, Rick Kalvoda, Ira Shaw, Nancy Wallace, John Worth, and especially Joe Nichols who contributed details on regulatory data sources and the FRY-14Q data. This is a "living" document which we expect to correct and update as we learn more.

1. Introduction

…the empirical background of economic science is definitely inadequate. Our knowledge of the relevant facts of economics is incomparably smaller than that commanded in physics at the time when the mathematization of that subject was achieved. Indeed, the decisive break which came in physics in the seventeenth century, specifically in the field of mechanics, was possible only because of previous developments in astronomy. It was backed by several millennia of systematic, scientific, astronomical observation, culminating in an observer of unparalleled caliber, Tycho de Brahe. Nothing of this sort has occurred in economic science. It would have been absurd in physics to expect Kepler and Newton without Tycho, and there is no reason to hope for an easier development in economics.

John von Neumann and Oskar Morgenstern
Theory of Games and Economic Behavior, 1944

The quote from von Neumann and Morgenstern's book, which inaugurated the influential field of game theory, sagely observed that, to have any hope of a lasting and meaningful impact on the world, a field of study requires the careful and comprehensive documentation of empirical facts. Though both renown as "pure" theorists, von Neumann and Morgenstern recognized that data is the ultimate arbiter in any enterprise that seeks to apply the scientific method. Less than two decades hence, with the introduction of the CRSP tapes, a data revolution began in the study of financial economics and, specifically, in the study of public equities and fixed income investments.[2] The availability of high quality and "clean" [3] data documenting historical individual stock performance has had a profound impact on financial economics and the financial-services industry. CRSP has since become the gold standard for high-quality data, providing both a long time-series history and large cross-section of firms (both of which are crucial for model validation and statistical hypothesis testing). This was subsequently followed by Compustat and other data sources – now seamlessly available via WRDS – that help integrate security returns and the underlying financial, operating, and ownership characteristics of those securities.[4]

The ensuing explosion in empirically driven research also coincided with the introduction of the ERISA standards in the 1970s, which impose certain fiduciary requirements on pension plans. These developments have fostered increasingly sophisticated practices among institutional investment

[2] The Center for Research in Security Prices (CRSP) was launched in 1960. Though eventually sponsored by a Ford Foundation grant, its beginnings were linked to the equity investment industry's needs for benchmarking itself against other asset classes.

[3] To provide usable stock return information the CRSP data must be adjusted for all manner of corporate events (such as stock splits and mergers). Best practices for eliminating outliers and non-representative entries emerged with time.

[4] Compustat, founded in 1962, a division of S&P Capital IQ, provides info on accounting data and offers integration with CRSP, Thomson I/B/E/S, management details (including compensation), insider & institutional holdings, analyst reports, credit ratings, and more. The Wharton Research Data Services (WRDS) offers the means to consolidate and integrate historical data from various providers to over 400 corporate, academic, government, and nonprofit institutions.

managers. The data revolution also had significant spillover effects into other capital assets, perhaps most notably in the market for derivative securities. While these contingent claims had existed for centuries, the virtuous combination of theoretical and empirical research has helped propel the notional amount of outstanding derivative securities to more than $500 trillion.[5]

Many might take issue with the statement that investment in equity, fixed income, currency, commodity, and derivative securities has been reduced to a science. Most of us, however, have been impacted by the rich interaction between academics who apply the scientific method and practitioners whose very livelihood rests on determining an optimal balance between applying their "art" versus the models and principles tested by academics using historical data. The market has attested to this important symbiotic relationship with the staggering proliferation of ideas, products and tools that are employed by the investment industry and whose origins can be traced to academia.[6]

Among the community of researchers and practitioners in commercial real estate (CRE), there seems to be a sense that the field has lagged behind other asset classes in benefitting from synergies and close cooperation between academic researchers and practitioners. Correspondingly, there has been a significantly slower rate of adoption of new methods, products, and tools amongst CRE investment professionals. What motivates the writing of this document is the conviction that the main culprit is the relative lack of CRE data availability, quality, and transparency. This is despite the fact that the CRE industry has made immense strides in the breadth and depth of various available data sources over the last two decades. The problems are particularly acute for real estate private closed-end funds, which have increasingly become the focus of many institutional investors (e.g., the rise in value-added and opportunistic fund offerings has been breathtaking over the last 20 or so years). Borrowing from von Neumann and Morgenstern, and liberally editing, one might say that it would have been absurd in mainstream finance to expect Black, Engle, Fama, Merton, Scholes, or Shiller without datasets like CRSP-Compustat, and there is no reason to hope for an easier development in CRE.

There are reasons for CRE professionals to seek data parity with peer asset classes other than the purely aspirational promise of a proliferation of new and validated methods, tools, and products. Over the past two decades institutional investors have experienced increased allocation to "alternative" private-market assets like CRE. This trend is widely expected to increase with the very recent addition of "Real Estate" as a distinct investment sector in the Global Industry Classification System. Indeed, CRE comprises a large asset class that is crudely estimated at $18 trillion in total value in the United States alone – this can be compared with roughly $30T of US corporate equity and $23T in household-owned US residential real estate.[7] Of the $18T, roughly $6 trillion is thought to be composed of "institutional

[5] Source: Bank of International Settlements.

[6] A limited list would include index funds, exchange-traded funds, volatility indices and derivatives, various swap markets, hedging strategies for a vast range of derivative products, and a litany of risk and performance measures.

[7] US CRE market value, current as of this writing, is calculated by marking to market the findings in Florance et. al., (2010) "Slicing, dicing, and scoping the size of the US commercial real estate market," *Journal of Real Estate Portfolio Management*, v. 16 (2), 101-118. The marking to market calculation is performed using the Moody's RCA Analytics National All-Property Commercial Property Price Index. The corporate equity and residential real estate figures are taken from the 2016Q3 release of the Federal Reserve Bank's Financial Accounts of the United States.

quality" assets.[8] The institutional quality CRE asset base in the U.S. is comparable in size with the Treasury bond market (prior to the recent financial crisis) or with the publicly traded corporate bond market. Institutional investors have traditionally been the primary users of the methods, tools, and products that resulted from the data revolution in mainstream finance. With increasing allocations to CRE, it seems reasonable to expect that CRE professionals working in institutional settings would face increasing pressures to deploy quantitative tools and transparency on par with the other components of their institution's portfolio.

Just as the fundamental securities in the context of equities and fixed income are stocks and bonds from which one may construct portfolios, the fundamental security in CRE is an individual property. In what follows, we will refer to a "comprehensive historical CRE dataset", by which we mean a representative and research-quality (i.e., "clean") CRE property-level dataset that spans two or more cycles and that is sufficiently rich in physical and financial details for most referenced properties to enable a casual pro forma analysis at various points in time. Similar comprehensive datasets exist for stocks and bonds (e.g., through WRDS). So why does CRE data lag behind other classes in availability, quality, and transparency? To understand this it is useful to first observe that:

1. Most properties are held by private institutions or individuals who are subject to few, if any, public reporting requirements.
2. Publicly traded and regulated entities that own CRE are seldom required to report granular data at the property level.
3. Real estate is highly illiquid and transactions, while often part of a public record, occur infrequently. Moreover, key variables necessary for the calculation of economic performance (e.g., capital expenditures and cash flow information) need not be reported.

Despite these systemic limitations, the labor and financing intensiveness of CRE asset management has spurred the existence of a host of service providers who, in the course of their business, collect or generate data at the property level. These include brokers (for sales or leasing), advisory and appraisal firms, professional management companies, and specialized accounting platforms. The value of the staggering amount of data amassed by this network has not been lost on market participants and there are by now numerous CRE data providers who curate an impressive variety of property-level information. In principle, the building blocks may exist for the creation of a comprehensive historical CRE dataset. The challenges for combining such building blocks, if they do indeed exist, are numerous:

1. A first challenge is the sparse nature of the data – either in the cross-sectional or in the time-series dimension. Perhaps because of the manner by which providers source their data, each dataset typically lacks some key set of fields and/or is not sufficiently representative of the universe of properties and/or does not track properties back far enough. Overcoming this would require linking or joining a variety of existing data sources. Such an exercise would be far from

[8] This assumes institutional quality assets comprise roughly a third of all CRE in the United States, consistent with Geltner, D., N. Miller, J. Clayton, and P. Eichholtz, (2013), "Commercial Real Estate Analysis and Investments." OnCourse Learning, 3rd edition.

trivial because there is no universal way by which properties are identified across data providers.

2. Even if the merging issues could be overcome, a second challenge concerns access to individual data sources. Specifically, many vendors are only willing to permit individual researchers access on a case-by-case basis and, at times, at high costs. This is despite the fact that historical data tends to be a marginal source of revenues for most providers. In other words, thus far the value proposition in contributing to a consortium of historical datasets (as exemplified by WRDS in the case of other asset classes) has not been recognized by key data providers.
3. Related to the previous point, data providers often harvest their data from participating owners or managers. A key concern is compromising the confidentiality of their data contributors (e.g., ownership, exact property location, attribution of performance, etc.). To overcome this, data providers who agree to work with academic researchers require the signing of a non-disclosure agreement (NDA). A hidden cost of this is the fact that every researcher granted access to a specific source must individually invest resources in cleaning raw data – there are therefore limited opportunities to benefit from the scale of prior research.

A key goal of this document is to assess and more fully flesh out the scope of the challenges listed above and, based on that exercise, propose possible approaches to overcoming them. It is important to emphasize that we consider this to be a living document which summarizes our limited and incomplete understanding of an inherently complex set of issues.

The paper proceeds as follows: In Section 2, we summarize an attempt to identify and roughly catalogue the data available from a number of commercial and non-commercial institutions. We do this in an attempt to learn whether key building blocks even exist that would allow for the creation of a comprehensive historical CRE dataset. In Section 3, we dive further into the three challenges described above, and follow with a discussion of possible approaches to address them. Section 4 describes other CRE data initiatives that are currently in progress. Section 5 outlines our own proposal for a data initiative based on what we have learned thus far. Section 6 Concludes.

2. Summary of available CRE data

Table 1 documents the list of entities whose CRE-related data we summarized. The individual summary details are available in Appendix A. We make no representation about the accuracy of data in any individual source, leaving that to future analysis. Different sources were surveyed at varying breadth of details. Moreover, a number of potential data sources are missing from the list below. All of these qualifications reflect the incomplete status of our efforts. Although we will make an attempt to discuss additional sources later in this section, we re-emphasize that this is a living document and subject to updating as our exploration evolves. Our main focus, for now, is on U.S. property-level CRE data. With these caveats in mind, we begin by classifying the various surveyed entities into four functional areas:

1. RE Services Providers – These are firms that provide services to property managers and investors. Such services include appraisal/valuation, brokerage and advisory, portfolio/property accounting and analysis software, and research/analysis.
2. Data Vendors – These are firms that collect and sell CRE data to various market participants.
3. Regulatory and Industry Sources – These are non-profit organizations that, in serving their regulatory mandate or their industry, collect and, in some cases, provide data. Detailed micro-level collections maintained by regulatory agencies are subject to significant restrictions on data access.[9]
4. Other – These are entities that collect or are in possession of CRE-related data. In particular, this classification is reserved for data sources that might not be "traditionally" considered by investors and other financial stakeholders in CRE. With the emerging importance of "big data" and an awareness of the interconnected nature of the CRE industry, we feel that this category is especially relevant to consider on a forward-looking basis.

RE Services Provider	Data Vendor	Regulatory and Industry Sources	Other
Altus	Axiometrics	NAIC	AirSage
Burgiss Associates	Buildfax	NCREIF	City energy (disclosure data LL84)
CBRE	CoStar	Federal Reserve (FRY-14Q)	
Greenstreet	MSCI / IPD		
JLL	RCA		
Preqin	REIS		
Stepstone	SNL		
	Trepp		
	Xceligent		

Table 1: List of surveyed data sources by function

We begin our analysis by breaking down the type of available property-level information into four broad categories: Physical characteristics, Economic performance, Lease information, and Capital stack. Each

[9] Data at most regulatory institutions may be used for external research only by staff at those institutions, with strict review of any publically released paper or presentation to insure that no confidential data is released.

of these, in turn, is further divided into subcategories of details that are key economic drivers for the asset. For each subcategory, Tables 2-5 document sources that may currently be accessed by academic researchers and that provide a high level of detail over at least one cycle.[10] Table 6 provides source coverage across geography, property types, and property class for all entities in Tables 2-5.

Physical Characteristics	Entities Providing Moderate to High Level of Detail
Class/quality	Axiometrics, BuildFax, CoStar, REIS, Xceligent
Size, age and Type	Altus, Axiometrics, BuildFax, CoStar, FRY-14Q, MSCI, NCREIF, RCA, REIS, Trepp, Xceligent
Construction	Axiometrics, BuildFax, CoStar, Xceligent
Floorplans	CoStar, Xceligent
Amenities	BuildFax, CoStar, MSCI, Xceligent

Table 2: Data sources that may be accessed by academic researchers and that provide a moderate to high level of detail for property physical characteristics over at least one cycle

Economic quantity	Entities Providing Moderate to High Level of Detail
Rental revenues	Altus, Axiometrics, FRY-14Q, MSCI, NCREIF, REIS, SNL, Trepp
Operating expenses	Altus, MSCI, NCREIF, REIS, Trepp
Capital expenditures	Altus, BuildFax, NCREIF, MSCI
Appraisals	Altus, NAIC, NCREIF, MSCI, Trepp
Transactions	Altus, CoStar, NCREIF, MSCI, RCA, REIS, SNL

Table 3: Data sources that may be accessed by academic researchers and that provide a moderate to high level of detail for property economic performance characteristics over at least one cycle

Lease variable	Entities Providing Moderate to High Level of Detail
Tenant	CoStar, MSCI, Xceligent, REIS, SNL, Trepp
Suite Information	Xceligent
Maturity	Altus, CoStar, MSCI, Trepp, Xceligent
Tenant allowance and concessions	Axiometrics, CoStar, Xceligent
Rent escalations	Xceligent

Table 4: Data sources that may be accessed by academic researchers and that provide a moderate to high level of detail for property lease structure and characteristics over at least one cycle

[10] The phrase "may currently be accessed by academic researchers" should be interpreted cautiously. It includes institutions that have been very open to providing data to researchers, those that are willing to sell data to researchers, those that only permit internal personnel to use data as part of a collaboration with academics, and those where senior executives have expressed an openness to the principle of access for academic purposes.

Economic quantity	Entities Providing Moderate to High Level of Detail
Mortgage debt	FRY-14Q, MSCI, NAIC, NCREIF, RCA, SNL, Trepp
Other debt	SNL
Owners, stakes and/or managers	Burgiss , CoStar, MSCI, NCREIF, Preqin, SNL , StepStone, Xceligent
Partnership terms	Burgiss, Preqin, SNL, StepStone

Table 5: Data sources that may be accessed by academic researchers and that provide a moderate to high level of detail for property economic performance characteristics over at least one cycle

Institution	Geography	Property Types	Property Class	Notes
Altus	Major MSAs	All major types	Core properties	Correspond to the ODCE subset of NCREIF
Axiometrics	500+ MSAs	Multifamily	All	Comprehensive within covered markets (capture most properties with 40+ units)
BuildFax	All markets	All	All	100% coverage for all population centers above 50k
Burgiss Associates	All markets	All major types	All	Reflects covered PE funds
CoStar	All markets	All	All	Arguably the greatest coverage
FRY-14Q	All markets (limited International)	All, including construction	All	Access restricted to internal staff, detailed information on financing terms and loan performance
MSCI / IPD	NA	All	NA	Estimated to represent 12% of properties held in professionally managed portfolios
NAIC	All markets	All	All	CRE collateral for insurance company debt
NCREIF	Major US MSAs	All	Institutional properties	Geography reflects locations of member firms' properties
Preqin	All markets	All major types	All	Reflects covered PE funds
RCA	All markets	All	All	Details for all property transactions over $2.5M
REIS	All major markets	All	All	Sparser coverage for smaller properties
SNL	All markets	All	All	Reflects public REIT ownership
Stepstone	All markets	All major types	All	Reflects covered PE funds
Trepp	All markets	All major types	All	CMBS properties
Xceligent	50+ MSAs	Office, Industrial, Retail	All	Some multifamily coverage with sparseness varying by market.

Table 6: Source coverage across geography, property types, and property class for all entities in Tables 2-5.

It is apparent from Tables 2-5 that no single data provider offers comprehensive coverage across the 19 items, with CoStar, MSCI and Xceligent offering moderate to high levels of detail on 10-11 of the items. As suggested in Table 6, the data coverage for each provider may not be comprehensive across geography, property types, or property class. Furthermore, several of the data providers track properties based on ownership. Specifically, to be tracked a property must be owned by clients for Altus and MSCI, member firms for NCREIF, and participating funds for Burgiss, Preqin, and Stepstone. If a property is sold by a client, member firm, or fund to a non-participating entity then it will no longer be tracked. These data sources, therefore, can generally only provide a partial history of the life of a property.[11] These issues, in a nutshell, capture the first set of primary challenges facing any attempt to create a comprehensive historical CRE dataset.[12]

We now turn to a selective discussion of various groupings of the data sources, some of which appear in Tables 1-6 and some not yet explored by us.

THE BROKERAGE FIRMS

CBRE, Cushman Wakefield, JLL, and Colliers oversee a vast network of CRE brokers (sales and leasing) and advisors (investment and property management). Through this network, these firms have amassed a large set of proprietary property-level data (especially on leases). The internally sourced information has generally been used to augment data purchased from vendors such as CoStar. In turn, the augmented data supports subsidiaries of the brokerage firms that offer all manner of market research to their clients. Before 2012, the major US brokerage firms had agreements with CoStar (from whom they purchased data services) restricting them as CoStar's clients from providing any CoStar competitor with CRE information (including internally sourced information). In 2012, as part of a merger settlement with the FTC, CoStar agreed to discontinue this practice.[13] This history, together with the past reliance of the major brokerage firms on augmented CoStar data, makes it difficult to obtain historical data directly from the brokerage firms.[14] As a result, despite the richness of information collected by the major brokerage firms, without appropriate permissions and cooperation from CoStar it is unlikely that these entities could significantly contribute to a comprehensive CRE historical database.

THE PRIVATE EQUITY SERVICE PROVIDERS

Burgiss, Preqin, and Stepstone provide investors in real estate private equity funds with advisory, market data, research, and/or investment management services. In that capacity these platforms have accumulated substantial amounts of data, some of which is at the property level. Burgiss and StepStone

[11] NCREIF's NPI-qualifying properties, for instance, have a median holding period of roughly seven years although this varies greatly by the acquisition year and other criteria.
[12] Another issue concerns the longitudinal depth of the coverage. Virtually all of the datasets considered in tables 2-6 have data going back to the late 1990's or early 2000's. This arguably comprises somewhere between one and two cycles. It is likely that by the end of this decade most of the sources considered will have amassed enough data to span two cycles.
[13] See https://goo.gl/T2AxLr.
[14] It would be challenging to go back into a historical database, founded on CoStar data and augmented using internal information, and be able to tease apart what information came from CoStar and what was internal.

acquire data from quarterly reports and due diligence documentation. Preqin obtains data through the Freedom of Information Act (as it applies to public pension and endowment funds) and general partner relationships (see summaries for more details). To our knowledge, the property-level data is not organized in a way that is easy to access or analyze (e.g., many of the reports are in PDF files of scanned print literature). All three platforms, however, have a successful track record of collaboration with academic researchers and there is much potential in harnessing the information they possess. An important advantage these data sources have, at least in theory, is the documentation of ownership structure and partnership terms. In particular, it is easy to conceive that their data would be indispensable in attempting to answer questions about the economics of the private real estate fund structure. Another potentially important role that the data in these platforms could play would be to shed light on riskier real estate assets (so-called core+, value-added, and opportunistic properties).

PROPERTY ACCOUNTING AND PORTFOLIO MANAGEMENT SOFTWARE FIRMS

Thanks to innovations in information technology, real estate property management has come to heavily rely on software tools that do everything from track property maintenance schedules to optimizing asking rents in a manner reminiscent of airline ticket pricing. This, of course, leads to massive amounts of collected data. In cases where the software vendor also provides data warehousing services the data is centralized. While we have yet to survey them, firms like Entrata, Yardi, and RealPage could potentially play an important role in enriching details of cash flow and capital expenditure data unavailable elsewhere.

PROPERTY APPRAISAL AND ADVISORY FIRMS

CRE properties are often subject to professional appraisals when undergoing a due diligence process, or in order to qualify for financing, or to satisfy the transparency requirements of various stakeholders. The appraisal process involves documenting a host of information about physical attributes as well as a history of leasing and cash flow components. While the data produced in the appraisal process may not typically be stored in a format that can be easily ported into a standardized template, the existence of the data holds some promise that the right resources and innovations in information technology may render this type of data usable in the future. It may be helpful that the number of firms that provide most of the appraisal services for institutional quality assets is not large.[15] The Altus Group, in particular, may hold the greatest amount of promise in this regard because it wholly owns Argus, a leading CRE valuation and cash flow modeling software firm. Altus is currently developing capabilities to link Argus Enterprise to their ODCE clients' analysis platform database.

CONFIDENTIAL SUPERVISORY DATA

Commercial banks – According to the Federal Reserve Board's 2016 release of its Financial Accounts of the United States, balance sheet CRE mortgage loans issued by banks (U.S. chartered depository institutions) comprise roughly half of all CRE mortgages. The banks are required to report aggregate statistics for such loans in their call reports. Detailed loan-level data, however, is not to our knowledge

[15] Some well-established appraisal firms include Altus Analytics, Integra Realty Resources, and Situs-RERC.

available in any systematic way save for three places: the banks themselves, the FDIC which insures depositors against bank failure, and the Federal Reserve which maintains detailed security holdings data for banks that must undergo stress tests. In the latter case, the data is considered highly confidential and access is strictly limited to Fed personnel with sufficient clearance.[16]

The FRY-14Q data collection that is surveyed in this document supports (1) the Federal Reserve's assessment of the capital adequacy of large bank holding companies (BHCs) using forward-looking projections of revenue and losses, (2) supervisory stress test models and (3) continuous monitoring efforts. The H.2 schedule of the FRY-14Q collection contains detailed data on credit facilities at participating institutions that are backed by CRE properties. Beyond the FRY-14Q collection, another source of confidential supervisory CRE data is the FDIC's Loss-Sharing Loan Data. The FDIC maintains data on loans where they have a loss sharing agreement (a strategy often used to dispose of assets held by failed banks). A significant share of such assets with loss sharing agreements tend to be CRE loans, though they do represent a very specific subset of the CRE market.

Access to confidential supervisory data is highly restricted, even within the agencies collecting the data. Non-agency researchers do have two alternatives. The first option is to collaborate with researchers at the regulatory agencies. Most agencies have formal policies governing such collaboration. The alternative is to directly negotiate with the participating institutions for access to the internal data used to construct the regulatory data submission. A wide range of influential research papers on CRE have been written in the past using similar proprietary data from a single firm, or a small group of firms. One advantage of using proprietary data tied to a supervisory collection is access to the publically available instructions for the collection prior to acquiring the data, significantly simplifying the initial data work for the researcher.

OTHER POTENTIAL SOURCES OF CRE DATA

We list and discuss these in no particular order:

1. Green Street Advisors – considered something of a gold standard in the quality of public REIT analysis and reports. Green Street mostly uses other platforms for their net asset value (NAV) REIT portfolio analyses (like SNL). They do maintain three internal databases: 1) Green Street's Mall Database, which comprises information and insights into over 1,300 regional malls and other retail properties across the U.S., with a primary focus on REIT-owned properties; 2) Green Street Advisors' Strip Center Database, which offers information on 2,500+ REIT-owned strip center properties across the U.S.; and 3) Green Street's Manhattan Office Database, which contains data on 100+ properties (including some important privately held properties) representing 120+ million square feet and an estimated USD 40 billion in value.
2. Municipal public records – As part of their tax and other laws, local governments record and collect a great deal of information about CRE. This includes tax assessments, building permits, zoning and easement information, and environmental/engineering data. In some cases this data is easily

[16] It is possible to connect, from public records of liens or from commercial datasets like RCA, bank loans and properties. Term sheet details, however, are not available.

accessible through a web portal and goes back two or more decades. In other cases, most of the data at best corresponds to scanned documents. There are firms that specialize in curating this type of data for commercial purposes. BuildFax, for instance, is a subsidiary of DMGI that compiles information on property condition based on mining public documents for building permits for CRE properties (with data stretching as far back as four decades). Another example is data, collected since 2010, from the NYC Mayor's Office of Sustainability corresponding to energy and water usage (in disclosure compliance with Local Law 84).

3. CMBS Prospectuses and Credit Company Reports – Annexes to CMBS Prospectuses filed with the SEC can contain a great deal of collateral-level information (especially for deals securitized after the Great Financial Crisis). These provide appraisal and cash flow information (which can also be found in the datasets of providers like Trepp). Likewise, credit rating companies have started to issue reports on CMBS deals that provide great details on the loans with the largest balance in the deal. Because the credit rating agencies are supposed to do their own due diligence on the collateral, these reports can contain complementary information to what already exists in the deal prospectus.
4. Insurers of CRE – Insurance companies seeking to underwrite CRE properties for various risks compile a great deal of descriptive information about the property's physical attributes and uses. To our knowledge, this is an untapped potential resource for researchers.
5. Unlike financial assets, CRE assets are part of the physical space that surrounds them and there is a sea of data that relates the properties to both the physical environments they create as well as the physical environments in which they operate. While it is beyond the scope of this exercise to attempt to list the endless types of variables that describe these relationships, it would be foolish not to acknowledge their potential in helping us to understand CRE value creation whether measured in financial or social terms. AirSage, for instance, tracks wireless signaling data from major carriers. This can be instrumental in modeling, evaluating and analyzing location, movement and flow of people and assets. One can also envision that social media data can help shed light on trends and uses for CRE-relevant characteristics.

3. Towards a Comprehensive Historical CRE Dataset

In this section, based on the observations of the previous section, we outline the major data and agency issues that pose challenges for the creation of a comprehensive historical dataset. We then outline some thoughts on how these issues might be progressively addressed.

MAJOR DATA ISSUES

Our survey of CRE data sources suggests that the *amount* of data in existence may not pose a major hurdle. Below is a list of key hurdles.

1. Merging Disparate sources – As noted in the previous section, even the most detailed sources do not come close to being comprehensive in covering all of the items in Tables 2-5. Thus the creation of a comprehensive dataset would necessitate the merging of multiple datasets. An

existing obstacle is the fact that there is no universal building identification protocol that can function as a natural unique key for merging diverse property-level data sources. At the heart of the problem is the fact that the notion of a single "property" can mean different things in different contexts. A set of buildings with different addresses might be considered a single property (i.e., a complex). This happens with certain securitization deals in Trepp, with certain member-owned properties in NCREIF, and certain transactions in RCA or CoStar (examples abound for other datasets). Moreover, because parts of a complex can be separately sold or securitized, data in one set might not be perfectly correlated with data in the other.

2. Data Consistency – A distinct set of data incongruities can also arise even if properties are perfectly matched. This is because underlying sources differ across (and sometimes within) datasets, and/or because there is no universal agreement on how to calculate certain data items. For instance, information coming from brokers versus owners/managers concerning a sale or a lease transaction might not coincide; accounting items such as NOI and CapEx may not be uniformly calculated across different stakeholders or platforms; property class and quality are subjective; many more examples abound.
3. Coverage gaps – save for the category of "Physical Characteristics" (see Table 2), there are potentially serious coverage gaps in most of the major categories defined by Tables 2-5. At first blush, there may seem to be good coverage for economic performance measures (Table 3) across several providers. This is somewhat deceiving. Altus, MSCI and NCREIF provide enough property-level information to calculate holding period returns or appraisal-based IRRs. However, Altus and NCREIF focus on an important but far from representative set of properties. MSCI, on the other hand, at this point in time restricts access to historical property-level data to internal personnel because of confidentiality agreements with managers and investors who provide the data. All three entities, importantly, only cover properties while they are held by clients or participating firms. Once a property is sold, coverage generally ceases save for the minority of cases in which the subject property is bought by another client or participating firm. This would introduce gaps in the history of properties in a dataset populated only using data from Altus, NCREIF, and/or MSCI. Restricting attention only to properties while they have coverage could be problematic in that it may create the potential for "ownership bias".[17] For instance, it may not be appropriate to extrapolate findings from research conducted on ODCE assets to core assets owned outside of the ODCE universe.

   Similarly, the coverage of leasing details that might be attributed to CoStar and Xceligent from Table 3 might also be misleading. The lease information may reflect listings rather than actual signed leases and rent revenues may correspond to estimates rather than actuals. It would be difficult to surmise whether this may or may not pose a problem for

[17] A well-known example from research into open-end equity mutual funds in the U.S. concludes that restricting analysis to mutual funds that are in existence introduces "survivorship bias" and consequently overstates mutual fund returns by anywhere from 50 to 100 basis points (see Elton, Gruber, and Blake, 1996, "Survivorship Bias and Mutual Fund Performance", in the *Review of Financial Studies*, Vol. 9, pp. 1097-1120.) Likewise, properties might be dropped from ODCE funds for economic reasons that might be correlated with their economic performance. Not adjusting or controlling for this possibility may misstate the economic benefits of owning these properties. In practice, researchers often ignore such issues. Ideally, we should at least be aware of them.

research without doing a benchmarking exercise for which actual verified lease information would be necessary.

Finally, to our knowledge there is little readily available information (even for paying customers) on individual mortgages on banks' balance sheets or on mezzanine/bridge loans tied to the equity in individual properties. Historical information might reside with the major brokerage firms who frequently intermediate mortgage loans but, as explained earlier, it would be challenging to make it accessible.

AGENCY ISSUES

Unlike public equities, for which the disclosure of cash flow variables is mandatory and economic quantities like prices and dividends are fairly transparent, details of CRE assets and their transactions are not typically subject to mandatory disclosure. Because private market participants, for competitive reasons, are reluctant to reveal details of their business, timely granular information is extremely valuable. This drives a thriving market in CRE information, supporting an ecosystem of providers.

Most providers are legitimately concerned about "giving up control" of their data, even if it is historical data designated solely for academic purposes. Contributing the data may undermine the provider's ability to profit from it or, worse yet, lead to a situation in which source confidentiality is breached. In particular, many data providers depend on a network of brokers/owners/managers for sourcing their data and have confidentiality agreements in place to protect individuals or entities that contribute data. Finally, no data provider wishes to be placed in a situation where their data is shown by academics to be problematic or inferior to that of a competitor.

FACING THE CHALLENGES

The issues that may be simplest to address concern the proprietary and confidential nature of the data. Many academic research institutions already have in place systems, platforms, and protocols for dealing with highly sensitive data. Important examples are provided by medical studies or those employing individual medical records and where confidentiality is mandated by Federal legislation (the HIPAA Privacy Rule), studies involving human subjects (who are protected by the so-called Federal "Common Rule" policy), and studies that fall under the Federal Food and Drug Administration's set of rules for protecting human subjects. Leveraging such existing systems to provide the necessary level of security and confidentiality of granular historical CRE data would be straightforward and would likely involve no or little fixed cost expenditures by the data contributor.[18]

[18] The fixed costs for installing and cleaning the raw data are usually borne by grants funding the hosting institution. An example of the type of secure protocols mentioned and currently in use at the University of North Carolina is available at: http://onlinelibrary.wiley.com/doi/10.1111/cts.12060/pdf. As another example of a secure platform in a CRE context, please consult the description of CUSP in the next Section.

To mitigate profit concerns, access could entail a user fee.[19] An alternative access model would keep raw data on the providers' servers and by use of a mapping allow an authorized user to simultaneously pull data from several providers to create a "pre-cleaned" instance of a merged set. This is more akin to the WRDS model, which necessarily imposes fewer security restrictions on use of the data. It is particularly worthwhile to note that most data providers generate little revenue from data that is four or more years old. It is therefore arguable that access to a historical dataset that has already been merged and cleaned would attract more academic users willing to pay access fees to data-contributing entities.

It is also important to mention that many data providers already sell or even give away to academics historical data containing confidential items with only a signed non-disclosure agreement as the security mechanism in place. It would be hard to envision that a clean, merged, and highly secure platform that charges user license fees would compromise data providers' interests any more than do current ad-hoc and inefficient arrangements.

Another concern mentioned earlier corresponds to adverse "reputational" impact of work that brings unfavorable attention to the data provider. This concern is harder to dismiss, but it is also a constant risk with *non-academic* users. We are not aware of any cases in which a data provider suffered substantial business loss as a result of an academic publication. We are, however, aware of instances where academics have found problems with a dataset and published their findings with the end result being that the data vendor addressed the problem (which, in the end, made everyone better off).[20]

In striving towards a comprehensive historical CRE dataset, it is likely that the more daunting task, by far, would be merging the disparate data sources and cleaning the resulting dataset (which would entail an evaluation of consistency across shared fields). It cannot be emphasized enough that such an exercise, if successful, would likely lead to substantial benefits to the data providers who could translate any success in merging historical datasets into a product that enables clients to merge *current* disparate datasets. Another considerable challenge, that would still remain after merging and cleaning, concerns the inevitable gaps in many properties' history. At that point it would fall to researchers to employ or even pioneer tools to deal with the unbalanced panel data. Here too, advances with a merged historical dataset would accrue to users of current data and therefore the data providers.

4. Concurrent Data Initiatives

A number of data initiatives, focused on themes similar to the ones discussed here, have recently emerged. We feel that it is important for such efforts to coordinate whenever feasible for both reasons of efficiency and because all such efforts tend to rely on the good will of a multitude of contributors who are at risk of "data request fatigue". Below are brief overviews.

[19] The fee-based user license agreement model corresponds to the manner in which many data providers already contract with academic institutions.

[20] See: Ince, O. S. and Porter, R. B. (2006), "Individual Equity Return Data From Thomson Datastream: Handle With Care!," *Journal of Financial Research*, 29: 463–479.

THE DEPARTMENT OF ENERGY'S DATALab

Interest in energy efficiency and sustainability, both in the public and private sector, has experienced tremendous growth over the last decade. The Forum for Sustainable and Responsible Investment estimates a (roughly) four-fold growth in "sustainable, responsible, and impact" (SRI) investments over the last ten years and that currently roughly $1 out of every $5 is invested in SRI-related assets. Academic research on energy efficiency, sustainability, and CRE is likewise growing.[21] To be able to effectively assess investment performance, one must grapple with the data issues discussed in this document. In recognition of that, the U.S. Department of Energy (DOE) has recently undertaken an effort to support the creation of a dataset of CRE properties that allows researchers to relate economic performance to investments in energy efficiency and sustainability.

Essentially, the goal is "…to develop new resources and infrastructure related to building performance and financial data, catalyzing further research on the relationship between sustainability, building performance, and financial benefits."[22] To address the confidentiality and security challenges outlined earlier, and to ensure that the data resides with an independent rather than government entity, the Lawrence Berkeley National Laboratory will be the data custodian. The data entity itself will be known as DATALab (Data Aggregation & Trend Analysis Laboratory). Data contributors will deal only with the Lab and researchers wishing to use the data will sign an NDA with the Lab and be subject to the Lab's security protocols for de-identification of results. As of this writing, the DOE is in the process of recruiting data contributors.[23]

MIT's REAL ESTATE INNOVATION LAB

The MIT Center for Real Estate has established The Real Estate Innovation (REI) Lab – a research and development initiative aimed at exploring innovation in the built environment. The Lab has five principal areas of research that span from big data for the built environment, exploring the cutting edge technologies from the urban and real estate tech sector in its Urban and Real Estate Tech Database as well as research and data on 21st Century dimensions of innovation real estate for economic growth. As a principal and long-term project the Lab has established the CRE Tech Database, aimed at supporting research and development that enables innovation in the built environment. Using cutting edge and "big data" sources, this aims to be a comprehensive database of characteristics that describe the built environment (design/physical, financial/economic, legal/political).

---

[21] Examples of work by members of CREDA include:
Fisher, Jeff and Gary Pivo (2011), "The Walkability Premium in Commercial Property Investing", *Real Estate Economics* 39(2), 185-219.
Fisher, Jeff and Gary Pivo (2010), "Income, Value and Returns in Socially Responsible Office Properties", *Journal of Real Estate Research* 32(3).
Constantine E. Kontokosta, "Modeling the energy retrofit decision in commercial office buildings", *Energy and Buildings*, 131(1), 1-20.
Fuerst, Franz, Constantine Kontokosta, and Patrick McAllister (2014), "Determinants of Green Building Adoption", *Environment and Planning B: Planning and Design*, 41, 551-570.
[22] The quote is taken from a DOE data request summary titled "High Performance Real Estate Research Initiative".
[23] Interested readers may contact Cindy Zhu (DOE Fellow, Cindy.Zhu@EE.DOE.Gov) for more information.

The CRE Tech Databases's initial focus is on New York City, with the objective to serve as an R&D laboratory on the future of data in real estate, cities and the built environment. As one of the most competitive cities in the world, many resources are spent on trying to track its dynamic changes across many metrics. For the past two years, the Lab's leadership has worked with providers in the city to collect and create multiple datasets to measure these dynamics. The CRE Tech Database presently includes ongoing contributions from 15 private sector firms and industry associations. Finally, REI has built from scratch its own unique dataset on Accelerators, Incubators, Innovation Districts, and related real estate innovation products across the globe. Another singular contribution of REI is the ongoing aggregation and integration of the disparate sources and types of data into the Database. The principal contribution of the database is to conduct academic research projects that enable data aggregators to come together in a secure private and not-for-profit academic environment; combine new and existing data sources; create an infrastructure for research and development in the commercial real estate sector; disentangle price dynamics in the most valuable real estate market in the world; and gain insights into the future of real estate data.

Center for Urban Science and Progress (CUSP) at New York University

Created by Professor Constantine E. Kontokosta, the CUSP Building Data Repository is a platform to acquire, integrate, and manage all manner of data on commercial and residential buildings in urban areas. The repository currently holds data for more than 60,000 buildings across 10 countries and 46 cities, accounting for over 8 billion square feet of space. The buildings included in the database tend to be larger (over 25,000 square feet) and cover major metropolitan areas. Data include actual energy consumption by energy source; size and bulk characteristics of the building; property type and use mix; occupancy characteristics, including occupant density, operating hours, number of residential units, number of bedrooms, etc.; physical characteristics such as construction type, building age, etc.; spatial identifiers; lot characteristics (lot size, lot shape, floor area ratio); and spatial characteristics (corner or inside lot, building adjacencies, etc.). These building and lot data are integrated with correlative data, including building violation records, building footprint shapefiles, localized weather, tax assessment data, and tenant types. Data are collected from a diverse range of sources provided by city/county/state governments, federal agencies, private real estate companies, private data providers, building technology companies, and non-profit organizations.

A critical, and non-trivial, aspect of the repository is the spatial matching and geo-referencing of building, neighborhood, and city-level data across temporal scales. In total, the feature space of the integrated dataset approaches more than 10,000,000 data points for each year of data. This does not account for interval meter data. The data cleaning and integration of management activities involve several steps, from spatial joins of the various energy, land use, and socioeconomic datasets collected; to normalizing and standardizing field names, data definitions, and spatial identifiers across cities and data providers; to developing standardized descriptive data analytics to profile and characterize energy use in each building, portfolio, and city.

The Building Data Repository is maintained in accordance with the CUSP Data Facility. The CUSP Data Facility (CDF) is a secure research data environment with datasets, tools, and expert staff to provide

research support services to students, faculty, and city agency employees. All data and Data Facility assets are controlled by groups management and are housed within CUSP's private local area network. Researchers must authenticate through a gateway server in order to access remote workspaces. The most highly restricted data environment has no internet access and must be accessed on-site.

DOE's UNIQUE BUILDING IDENTIFICATION PROJECT

Independently of the DOE's DATALab initiative, the DOE's Building Technologies Office, with support from the Pacific Northwest National Laboratory, has recently launched an initiative to explore the creation of a Unique Building Identification (UBID) protocol. A pilot is expected to be underway by Summer, 2017. This effort, extending to both residential and commercial properties, is motivated by the difficulties in merging various disparate datasets. The hope is that an effective system, if widely adopted, would make easier the referencing of properties by federal and municipal agencies (tax, emergency, health and human services, etc.). If widely adopted in the public sector, one anticipates increasing adoption by the private sector. This might have an important (albeit *future*) impact on how data providers will organize their property-level information and may subsequently make the merging of diverse data sources easier.[24]

5. CREDA Pilot Study Proposal

At a meeting hosted by the University of North Carolina (October 20-21, 2016) and including academics, data provider executives, portfolio managers, investors, and Federal Reserve personnel, there appeared to be consensus that it would be immensely valuable to *all* if a mapping between various existing property-level data sources could be developed. Such a mapping would bring to bear various approaches and methodologies to effectively merge a disparate set of CRE data sources. In the process, the mapping would inform any attempt to create a Universal Property ID.[25] While researchers would still need permission to use the various databases, the mapping would open up the potential for many new avenues of research that were difficult if not impossible to do previously. To explore the feasibility of the mapping, it was suggested that CREDA would undertake a small-scale pilot study along the following lines:

Phase 1 (Target completion date of June 30, 2017): Identify a small sample (roughly 50) of properties about which a variety of willing data providers would likely have information. The goal would be to

1. Assess the level of consistency across diverse data sources in cases where there is overlap
2. Assess the feasibility of merging diverse data sources

[24] See http://buildingid.pnnl.gov/pdf/Buildings_ID_flier-v5.pdf for more information.

[25] This exercise would be complementary to the DOE's UBID project as it would focus on tying *historical* information about a property into a system of universal IDs. The model envisioned would likely function more similarly to the CRSP PERMNO referencing system. The latter tracks how securities are treated and linked following events like mergers, splits, or liquidations.

3. Devise and test approaches to a Unique Universal Property ID that can be used to refer to the same property across various data sources, and in the process propose how properties from each candidate data source can be mapped to this ID
4. Identify a set of important research questions to be explored in Phase 2.

Phase 2 (Target completion date of June 30, 2018): Based on what is learned from Phase 1, and subject to the continued cooperation of data providers from Phase I, the exercise would be scaled and applied to include 500-1000 properties, sufficient for actual research. In the process, it is hoped that two or more specific research projects would be launched that will utilize the pilot dataset and assess, in conjunction with the data contributors, the benefits of launching a larger/open-ended initiative.

To address the important confidentiality and security issues, all data contributed to the project would reside on a secure platform that is HIPAA-compliant, and only approved researchers will have access to the data. A list of individuals with access to the data will be available and all such individuals would sign non-disclosure agreements. Moreover, all output made available from or during the exercise would be reviewed to ensure that no property-level or otherwise sensitive information is released or resides anywhere other than the secure server. Although UNC's Institute for Private Capital (http://uncipc.org/) has agreed to be the sponsoring institution for the project, it is anticipated that researchers from other academic institutions, including CREDA members, will participate in the project. Proposals for partnering in Phase 1 have already been submitted to a number of data providers.

6. Concluding Remarks

There appears to be consensus for the view that Commercial Real Estate (CRE) lags behind other major asset classes in the quantity and quality of tools, methods, and products. Moreover, unlike other major asset classes, CRE has not experienced the high degree of integration between the investment community and academia. Our view is that a main (if not *the* main) reason for this is the relative absence of a high quality comprehensive historical dataset that can enable researchers and practitioners, alike, to study and validate models, tools, and methodologies across cycles and property characteristics. This document surveys some existing sources of CRE data and attempts to identify the key issues that face any attempt to move towards a researcher-accessible and comprehensive historical CRE dataset. The main challenges concern:

1. The difficulty in merging time series panels of disparate datasets that do not share the same set of property identifiers.
2. Consistency of similar quantities across disparate datasets
3. Coverage gaps in creating a time series history of any single property's attributes
4. Data provider concerns over losing profits from historical data
5. Data provider concerns over loss of confidentiality with respect to granular data

We consider potential solutions to address these challenges. Dealing with the merging issue might involve a sequence of studies with the collaboration of multiple data providers. The ultimate objective

would be to generate (1) a scalable mapping that would enable the merging of providers' historical datasets, and (2) a suite of tools that can be deployed to "clean" the resulting merged dataset by assessing/addressing consistency across shared fields and gaps in property history. An important promise of this approach is that it stands to benefit data providers and their clients because a successful outcome resulting from the exercise on historical data can be applied to *current* and therefore timely practitioner-oriented data. This would more easily allow clients to merge several platforms, increasing the appeal of any single dataset. Profitability concerns might be addressed by fee-based user licenses similar to what already exists with financial data (e.g., the Wharton Research Data Services). Finally, to address the confidentiality concerns, we point to successfully existing protocols and platforms in the social and life sciences where research is mandated by law to protect sensitive information.

Appendix A: Dataset Surveys

Item	Response
Data Provider	**American Council of Life Insurers**
Public contact	Khari Cook: KhariCook@acli.com, 202-624-2133 and Matt Wellens: MattWellens@acli.com
Type of Service	The ACLI is a Washington, D.C.-based trade association with approximately 300 member companies operating in the United States and abroad. ACLI's Commercial Mortgage Commitments (CMM) report provides information on the mortgage committments of life insurance companies, including property type, contract interest rate, debt coverage ratio, loan-to-value ratio, capitalization rate, and loan maturity. The data are based on surveys of 25-30 life insurance companies.
Short Description of Data	The ACLI Commercial Mortgage Commitments - Historical Database contains quarterly data for 10 different property types and in total. The following property types have been consistently reported since 1965: Apartments, Office Buildings, Retail, Industrial, and Hotel/Motel. The survey includes long-term (over one year) mortgage commitments on commercial properties in the US, including maturing balloon mortgages which have been refinanced for more than one year at current market terms. It excludes construction loans without the permanent mortgage financing, standby loans, loans secured by land only, social responsibility loans, tax-exempt loans, purchases of existing mortgages and acquisitions of mortgage-backed securities. All loan data are aggregated and distributed on a quarterly basis five to six weeks after the end of the quarter. All reported data are dollar weighted (by size of mortgage committmen) except for average loan size which is based on a simple average.
Geography	The ACLI Commercial Mortgage Commitments - Historical Database is distributed in an Excel file with all data aggregated to the national level. Their documentation reports that they capture the city, county, and state location of each loan (presumably it is the location of the loan collateral).
Frequency and Dates of Coverage	Quarterly since 1965 for all loans covered. Separate data for fixed rate loans is generally available since Q1 2000.
Underlying Data Source	Responses to a quarterly survey of loan commitments made by life insurance companies. All loan data are aggregated. Reported data are dollar weighted (by size of mortgage committmen) except for average loan size which is based on a simple average.
Completeness of Panel	
Physical Property Characteristics	Their documentation reports that they capture the rentable area of non-apartments and non-hotels.
Economic Performance Data	No performance data.
Ownership and Financial Structure Data	The survey responses pertain to loans originated by life insurance companies to CRE investors who are either purchasing or refinancing an existing property.
Caveats	Life insurance loan to CREd investors represent a relatively smaill perentage of total CRE lending.
Research Use	The stated cost for the CMC Historical Database is $2,125. Quarterly updates are $250, although one year (4 quarters) can be purchased for $500.
Academic Research Papers Using These Data	Fisher, Jeffrey D., David M. Geltner, and R. Brian Webb. "Value indices of commercial real estate: a comparison of index construction methods." The journal of real estate finance and economics 9.2 (1994): 137-164.

Item	Response
Data Provider	**AirSage, Inc.**
Internal Contact	
Public contact	(404) 809 2499; rkinskey@airsage.com
Type of Service	AirSage is one of leading location data provider in the U.S. It collects and analyzes wireless signalling data from major wireless carriers and turn it into meaningfull and actionable insight. The company provides this aggregated location information to various organizations such as private business, goverment agencies that are interested in modelling, evaluating and analyzing location, movement and flow of people and assets.
Short Description of Data	AirSage data receives wireless signals from carrier providers based on activity of mobile devices operating on cellular network, such as calls, text massages and data transfers). It generates over 15 billion phone location data per day, with average 100 data points per device per day. The average location accuracy is within 300 meters range (often within 50 - 100 meters). The data is anonymized and ingested to patented Wireless Signal Extraction (WiSE™) platform where is aggregated, analyzed and extracted in CSV format. The analyses includes aggregartions, time of day, resident classes, trip purpose, site analytics etc.
Geography	U.S. where cellular network coverage is available
Frequency and Dates of Coverage	real-time or near real-time; historical data is available since January 2009
Underlying Data Source	Mobile device users activity data from nationwide wireless carrier partners.
Completeness of Panel	According to the company, the analyzed data comes on average from more than 100 million devices each day.
Physical Property Characteristics	N/A
Economic Performance Data	N/A
Ownership and Financial Structure Data	N/A
Caveats	Data available at various spatial and temporal aggregations
Research Use	Data available for purchase; reduced pricing for research use
Academic Research Papers Using These Data	

Item	Response
Data Provider	**Altus (Altus Analytics)**
Public contact	info@altusgroup.com
Type of Service	Altus Group, which owns Argus and Voyanta, provides a variety of support services to the CRE industry. These include, valuation/appraisal, software & data solutions, and tax/cost/project management consulting. In late 2014 Altus & NCREIF launched a data platform for performance benchmarking and attribution against up to 22 of the 24 NCREIF Open-end Diversified Core Equity (ODCE) funds. The platform combines ODCE fund data from NCREIF and property-level data from Altus's clients.
Short Description of Data	Similar to NCREIF: Detailed income, expense and capex data along with quarterly appraised values or transaction prices for (mostly) core institutionally owned CRE. Key data not available in NCREIF includes detailed valuation inputs and leasing metrics.
Geography	US
Frequency and Dates of Coverage	Quarterly going back to 1999 (beginning in 2008 the data is more robust and includes most of the NCREIF ODCE funds). Benchmarking data goes back to 2009Q4.
Underlying Data Source	NCREIF ODCE data and client's ARGUS-based valuation/analysis data. More than 90% of the valuations are based on ARGUS-type model so all of the information going into these cash flow models are in principle available in some (potentially hard to use) format. Altus is currently developing a way to link ARGUS Enterprise to their platform database. Data is collected starting when they become an Altus valuation management client, although data is frequently backfilled for new clients for 1-3 years before engagement. When sold, data is no longer collected unless purchased by another Altus client (assigned a new unique ID but linked to prior ID).
Completeness of Panel	Clientele (and therefore property-level data) includes over 50 institutional clients, over 80 funds, over 6000 properties, over $450B in gross asset value. Includes, as a subset, over 90% of the NCREIF ODCE funds (by value). Representativeness of asset coverage is similar to NCREIF.
Physical Property Characteristics	Similar to NCREIF. In addition, includes geocodes, photos, property tax info.
Economic Performance Data	Frequency of tenant expiry by term (e.g., leases expiring over next three years), avg remaining lease term, avg rent psqft & per unit, market rent psqft & per unit, operating expenses psqft & per unit and growth rate, absorption/occupancy. Sales & partial sales, appraisals, key valuation metrics (cap rates, discount rates, terminal cap rates, NOI, market rent growth, PV of reversion value, expenses psqft, expense growth, CapEx data).
Ownership and Financial Structure Data	Owning institution/fund, fund type, manager, JV share of ownership. Debt and mortgage information is provided at an aggregate level similar to NCREIF (both property- and fund-level debt), but loan-level information is currently not available.
Caveats	None noted
Research Use	Not currently in use for academic research. Client benchmarking requires aggregation of at least three funds and five properties.
Academic Research Papers Using These Data	NA

Item	Response
Data Provider	**Axiometrics**
Public contact	http://www.axiometrics.com/company/contact
Type of Service	Exclusively *conventional* multifamily and student housing (no senior housing). Serve all the MF REITs.
Short Description of Data	Detailed income, expense and capex data along with quarterly appraised values or transaction prices for mostly core tax-exempt institutionally owned CRE.
Geography	over 500 markets (MSAs)
Frequency and Dates of Coverage	Back to at most 1995 (depends on the market). Every property gets updated once a month.
Underlying Data Source	Monthly active update, reaching out to owners/managers.
Completeness of Panel	Track anything that is 40 units or higher (soft rule). General criteria is client driven (often based on client request or deemed market relevant for comps or other metrics). Forecast in 140 markets - focus is driven by market interest.
Physical Property Characteristics	Floors, class of property, submarket grad, unit mix, sqftg. No amenity info (studnt housing description captures pet fees).
Economic Performance Data	Limited transaction info (very recent). Track rents occupancy, & concessions (for leases in that month). No actual rental revenues but can estimate them. Don't track expenses. Don't track CapEx. Student housing quantities are tracked by the "bed". Construction: pipeline, props coming on line (i.e., what's under construction); details on the construction deal such as #units, GP, key dates, ETA on market, owner, architect, status (leaseup, under construction). For leaseup: occupancy of units available (absorption), rents, concessions.
Ownership and Financial Structure Data	Manager, title owner(s), no debt, lender (incomplete along with some info on loan details).
Caveats	Generally very hard to get actual expense info! National Apartment Association has a survey that captures this in aggregate.
Research Use	
Academic Research Papers Using These Data	Very little at this point

Item	Response
Data Provider	**BuildFax**
Public contact	https://www.buildfax.com/public/contact/
Type of Service	For most clients: Property condition on select properties (e.g., roof age), amentities, systems (plumbing & electric), remodeling, and more. Also leverage data to provide market intelligence on construction and building permits.
Short Description of Data	Roof age, elec system age, plumbing system age, property condition characeristics. Demolished? Renovated? Not cosmetic changes. Pretty much everything that requires permitting but not necessarily the quality (spend amounts + type of change are noted). Always know: What happened (e.g., solar panels were installed), where it happened and when. Usually know a lot of additional details (contractor + cost). What determines the depth of the latter is jurisdictional practices.
Geography	All 50 states. Once information is documented, BuildFax has it. Focus is on population centers (70% of population). The more rural the less comprehensive the coverage. 100% coverage on every city over 50k. Substantial coverage for population centers above 25k.
Frequency and Dates of Coverage	On average goes back 21 years. As far back as 40 years and no shorter than 5 years. Updating is monthly for about 45% of the data, and the remainder is annual.
Underlying Data Source	US Building Permit Records + internal intelligence. Data is mined from text in the records.
Completeness of Panel	Cover single family homes + CRE. About 17-18% of dataset is CRE; about 2-3% is mixed-use. Anything residential rental property up to a 4-plex is considered resi. Otherwise classified as mixed-use or commercial.
Physical Property Characteristics	See above
Economic Performance Data	NA
Ownership and Financial Structure Data	NA
Caveats	
Research Use	No well-known track record thus far.
Academic Research Papers Using These Data	"Consumers and the Economy, Part II: Household Debt and the Weak U.S. Recovery", by Atif Mian and Amri Sufi, FRBSF ECONOMIC LETTER, January 18, 2011.

Item	Response
Data Provider	**Burgiss**
Public contact	research@burgiss.com, 111 River Street, 10th Floor, Hoboken, NJ 07030 USA. (201) 427-9600
Type of Service	Burgiss is a global provider of investment decision support tools for the private capital market.
Short Description of Data	The Burgiss Manager Universe is a research-quality database that includes the complete transactional (cash flows and quarterly NAVs) history for over 6,700 private capital funds with a total capitalization representing over $4.6 trillion in committed capital across the full spectrum of private capital strategies (as of June 2016). It is representative of actual investor experience because the data is sourced exclusively from limited partners, which avoids the natural biases introduced by sourcing data from general partners. The data includes the actual dates of the cash flows and is further supplemented with fund profiles. Access is available through the Private Equity Research Consortium and requires signing a non-disclosure agreement. Project analysis is conducted by a dedicated research assistant located at Burgiss offices and results of the analysis are passed to the academic researcher. Only aggregated results or output from statistical estimations are shared.
Geography	As of 2016-Q1: Burgiss maintained information on over 1,000 closed-end Private Capital Real Estate Funds and Funds-of-Funds. The Real Estate Funds portion of the dataset is further broken out by Opportunistic and Value-Added strategies, as well as Geographic Focus as follows: 684 (71.4%) in North America, 150 (15.7%) in Europe, 79 (8.2%) in Asia, and 45 (4.7%) in other regions or unkown.
Frequency and Dates of	Daily precision cash flows, quarterly NAVs, 1980-present
Underlying Data Source	Data is sourced exclusively from limited partners (LPs) and includes the complete, checkbook-style transactional activity (cash flows and quarterly NAVs) between LPs and their fund investments. This also allows for cross-validation when multiple investors are in the same fund.
Completeness of Panel	Estimate between 20-30% coverage of RE PE space, and increasing percentage coverage over time. Historical database includes 32 funds with 1980-1989 vintage, 115 funds with 1990-1999 vintage, 472 funds with 2000-2009 vintage, and 339 funds with 2010-2016 vintage.
Physical Property Characteristics	Limited information on specific properties is available from fund quarterly reports. These data are not yet available in electronic format.
Economic Performance Data	Daily precision cash flow information from which various measure can be built, e.g. IRR, DPI, etc.
Ownership and Financial	Some information is available at the fund level.
Caveats	
Research Use	Arms length access via a dedicated (embedded) research assistant. Researchers get dummy dataset(s) and supply STATA code for analysis.
Academic Research Papers Using These Data	Fisher, Lynn M., and David J. Hartzell, 2016, Class Differences in Real Estate Private Equity Performance, Journal of Real Estate Finance & Economics, p327-346.
	Falkenbach, Heidi, and Sami Kiehela, 2015, Performance of Non-core Private Equity Real Estate Funds: A European View, Journal of Portfolio Management 41(5), pp. 62-72.
	Fisher, Lynn M., and David J. Hartzell, 2013, Real Estate Private Equity Performance: A New Look. UNC working paper.

Item	Response
Data Provider	**CBRE**
Public contact	https://www.cbre-ea.com or www.erix.cbre.eu
Type of Service	CBRE provides a global comprehensive database with global coverage to deliver meaningful insights in to trends in CRE markets. The focus of CBRE data is real estate basics such as rents, cap rates, gross and net absorption and gross and net completions. With over 70,000 employees covering most markets they have the ability to augment standard dataset with high quality local intel.
Short Description of Data	CBRE"s coverage is global and comprises two major type of databases such as ERIX and Econometric Advisors (EA). ERIX is the database which contains detailed city by city information for major EMEA and Asian Pacific cities on office, retail and industrial real estate markets. The data are generally available from the early 90's with the exception from numerous markets due to market coverage. The EA database covers data for U.S. and Canadian major MSA's for the Office, Retail, Industrial, and Multifamily sector. Recently more alternative sectors such as hotels are available. The data is available on different levels according to the EA platform. Both platforms deliver forecasts and interpretations for the major cities. These datasets aggregate more granular data that is only available internally.
Geography	Global
Frequency and Dates of Coverage	Quarterly data for different time periods (depending on the city)
Underlying Data Source	Asia Pacific and EMEA data is updated by local research teams on a quarterly basis. The data goes through a quality control process and released to clients on the last working day of the month after the end of the quarter. In Americas, EA and research work together in compiling and validating data which is released throughout the EA web platform. Some EA data employs external sources such as Dodge
Completeness of Panel	-
Physical Property Characteristics	-
Economic Performance Data	Rents, Capital Value, Yield, Take-up, Vacancy, Stock, Availability, Completions, Pipeline, Absorption and various other variables.
Ownership and Financial Structure Data	-
Caveats	CBRE rent, cap rate and capital value data is mostly appraisal based with some exceptions eg transaction based data in the Torto-Wheaton index (EA). In addition, due to the global coverage, methodology and definitions can differ when compiling our variables.
Research Use	CBRE has a history of working closely with academics in providing data for scientific research. A non-disclosure agreement (NDA) must be signed by the researcher.
Academic Research Papers Using These Data	Chervachidze, S., and Wheaton, W. (2013). What Determined the Great Cap Rate Compression of 2000-2007, and the Dramatic Reversal During the 2008-2009 Financial Crisis?. Journal of Real Estate Finance and Economics, 46(2), 208-231. Crosby, N., Devaney, S., and Law, CV. (2012). Rental deprecation and capital expenditure in the UK commercial real estate market, 1993-2009. Journal of property research, 29(3), 227-246. Hendershott, P.H., Jennen, M. & MacGregor, B.D. (2013). Modeling Space Market Dynamics: An illustration Using Panel Data for US Retail. Journal of Real Estate Finane Economics, doi:10.1007/s11146-013-9426-z Sivitanides, P.S., Torto, R.G., and Wheaton, W.C. (2003). Real estate Market Fundamentals and Asset Pricing. *Journal of Portfolio Management*, 29(5), 45-53.

Item	Response
Data Provider	**CoStar**
Public contact	sales@costar.com; 1-888-226-7404
Type of Service	CoStar claims to have over 1,200 researchers and more than 130 field research vehicles collecting and verifying information from a variety of sources. Their research process includes thousands of calls daily to brokers and owners, reviewing court filings, tax assessor records and deeds, tenant canvassing, third-party data feeds and automated data collection of thousands of broker and retailer websites. Their field researchers inspect over 2 million properties annually and have taken over 12 million photographs. As of September 2016, CoStar claimed to have information on over 4.2 million properties constituting more than 95 billion in total square feet.
Short Description of Data	CoStar "Suite" is the combination of three CoStar products: CoStar Property, CoStar COMPS, and CoStar Tenant. CoStar COMPs is a database of CRE transactions. According to CoStar the COMPS database contains over 1 million verified sales records, across all property types. The data generally are available from 1997 onward. These cover at least 100 major MSAs. Each transaction record contains approximately 200 data fields. These fields include the transaction price, the sale date, and a large number of hedonic characteristics on the property. These include the property's address and zip code and latitude and longitude. In addition, a number of characteristics of the transaction are included, such as the name of the buyer and the seller, the name of the buyer's broker (if applicable) and the name of the seller's broker. However, not all data fields for each transaction are populated. The remainder of this description applies only to CoStar COMPs. The tenant database is said by CoStar to cover every CRE property in the major metropolitan areas. This building level information includes asking rents and vacancies, as well as property hedonics. More research is needed on the Property database.
Geography	Covers the entire US but coverage is concentrated in the major MSAs.
Frequency and Dates of Coverage	The transaction database is continuously updated.
Underlying Data Source	The transaction data come from calls to brokers and owners, court filings, tax assessor records and deeds, tenant canvassing, third-party data feeds and automated data collection of thousands of broker and retailer websites. The source of the transaction is not reported, although CoStar claims that all the data is "verified."
Completeness of Panel	CoStar claims to cover the universe of CRE transactions in the US. Unlike RCA, there is no minimum transaction price for inclusion in the database. However, CoStar has grown over the yeas by buying competing companies and adding their data to COMPS. This means that as you go back further in time, the coverage is less extensive. However, data for about the 50 largest MSAs is available back to about 1997.
Physical Property Characteristics	Each sale transaction contains approximately 200 data fields with information on the property, its location, and to a lesser extent, characteristics of the transaction, such as buyer and seller names. Property characteristics include: sale price, age of the property, square footage of the structure, land square footage, number of bedroom (if multifamily), the "class" of the property (A, B, or C), condition of the property, type of construction (metal, masonry, brick, etc.), and the assessed (taxable) value of the property that is useful in some research.
Economic Performance Data	No performance data.
Ownership and Financial Structure Data	
Caveats	There seems to be variation in the prices they quote different academic institutions. Their daily download limits make it difficult to create large databases.
Research Use	The University of Florida just purchased access to the COMPS data for $24,000 per year. However, CoStar places a limit on daily downloads--typically 500 transactions per day. This makes the construction of a research database quite tedious.
Academic Research Papers Using These Data	Florance, Andrew, et al. "Slicing, dicing, and scoping the size of the US commercial real estate market." Journal of Real Estate Portfolio Management 16.2 (2010): 101-118. Nichols, Joseph B., Stephen D. Oliner, and Michael R. Mulhall (2013). "Swings in Commercial and Residential Land Prices in the United States," Journal of Urban Economics, vol. 73, no. 1, pp. 57-76.

Item	Response
Data Provider	**Federal Reserve System**
External contact	https://www.federalreserve.gov/apps/reportforms/reportdetail.aspx?sOoYJ+5BzDZGWnsSjRJKDwRxOb5Kb1hL
Type of Service	The FR Y-14 collects detailed data on bank holding companies' (BHC) various asset classes, capital components, and categories of pre-provision net revenue (PPNR) on a quarterly basis. Schedule FR Y-14Q H.2 is a quarterly schedule that collects information on credit facilities containing CRE loans. Given that the majority of the credit facilities contain single loans, this document will use the terms credit facility and loan interchangeably.
Short Description of Data	The data collected includes information on loan status, underwriting terms, pricing, type of loan, and characteristics of the property both at origination and as-of the reporting period. The database includes multi-family loans, loans on Non-Farm, Non-Residential properties, construction loans (including single-family construction loans), and loans secured by CRE originated in non-domestic offices.
Geography	BHCs with more than $50 billion in assets report in the collection information on all credit facilities with committed values of over $1 million, regardless of geographical location. This includes loans made by participating institutions secured by CRE properties in other countries. While the lending footprint of the participating institutions does span all 50 U.S. States, the lower limit on the total commitment at the credit facility level combined with the lower limit on institution size does result in less coverage in areas with low CRE property prices.
Frequency and Dates of Coverage	The initial collection began in June of 2012 with 19 of the largest firms. The collection was extended in 2014 to include all BHCs with more than $50 billion in consolidated assets. The collection also will eventually include firms other than BHCs identified by the Federal Stability Oversight Council as systemically important as well as IHCs formed to consolidate the holdings of foreign banking organizations. The collection is a quarterly panel.
Underlying Data Source	The FR Y-14 schedules is an official regulatory collection. The participating firms provide the FR Y-14Q Schedule H.2 data to the Federal Reserve on a quarterly basis. The collection includes a series of edit checks on the completeness of the collection. Federal Reserve staff work closely with participating institutions to address issues with the quality or completeness of the data reporting.
Completeness of Panel	As of 2014:Q3 FR Y-14 respondents held $13.9 trillion in assets, accounting for 70 percent of the assets held by all BHCs, savings and loan holding companies, and commercial and savings banks, based on the total reported assets in the Call Report and Y-9C. The coverage of total CRE loans is less than the share of total assets, primarily due to the disproportionate role of smaller banks in CRE lending. Among the Y-14 respondents there are two additional gaps in the coverage. 1) Firms with non-material CRE portfolios, currently defined as less than $5 billion and less than 5% of tier 1 capital, are not required to report their CRE loans on the FR Y-14Q H.2 schedule. 2) Firms are only required to report credit facilities whose committed balance is greater than or equal to $1 million.
Physical Property Characteristics	The database contains the property type of the building (office, retail, industrial, multi-family, and hotel). The database contains the zip code of the building, but not the specific address. The age and size of the building is also included. Size is reported in units for hotels and multi-family, and square feet for other property types. In cases where there are multiple properties securing the loans in the credit facility, these fields are completed based on the largest property, or property type, in the collateral pool. The purpose of construction and land loans is also given, including the property type if the project is for commercial real estate. Single-family construction loans are classified as CRE loans, and are include in the collection.

Economic Performance Data	The property value and annual NOI at origination are reported. The database also collects the current value, NOI, and occupancy rate - along with the date at which each of those values were last updated. BHCs report for each credit facility the current loan status, including days past due and the non-accrual status. The collection includes the internal risk rating the BHC has currently assigned to the loan, as well as the corresponding external rating based on BHC supplied concordance tables. The share of the loan that has been charged off and the level of the loan-specific reserves held by the BHC against the specific credit facility is also reported.
Ownership and Financial Structure Data	The underwriting terms and pricing for the loans are included. The underwriting terms are provided, or can be derived, both at origination and based on current reported values. The current outstanding balance is reported, indicating the degree of draw on construction loans and lines of credit secured by CRE properties. The name of the property owner is also provided, unless the owner is an individual. In that case the identity is not provided.
Caveats	Schedule H.2 of the FR Y-14Q collection does not include information on loans secured by owner-occupied CRE. Information on those loans are collected on Schedule H.1 with other loans to corporations and businesses. That schedule collects information on obligor financial conditions, instead of characteristics of the property used as collateral. The FR Y-14 is an official regulatory collection. The Federal Reserve annually proposes revisions to the collection that are designed to either improve information on the risk characteristics of the portfolios or to reduce reporting burden. This process includes an official public call for comment on the proposed changes. The nature and structure of this collection may change over time, due to this process.
Research Use	The FR Y-14 data has been identified as Confidential Supervisory Information (CSI) data within the Federal Reserve System. The data is not and will not ever be made publically available by the Federal Reserve. Federal Reserve staff may request access to the data to support external research projects. Such projects should not include any results that would allow a reader, either based on the published results or in conjunction with other public or private data, identify a specific BHC or obligor. Federal Reserve staff may work with external co-authors, but the external co-authors will not be provided access to the FR Y-14 data. Individual BHCs in the FR Y-14 data collection have complete control over the data in their submissions, and have the right to negotiate access to that data with researchers and consultants.
Academic Research Papers Using These Data	Black, L., Krainer, J., & Nichols, J. (2016). From origination to renegotiation: A comparison of portfolio and securitized commercial real estate loans. *The Journal of Real Estate Finance and Economics* , 1-31.

Item	Response
Data Provider	**Green Street Advisors**
External contact	www.greenstreetadvisors.com
Type of Service	North American and European REIT research database
Short Description of Data	Green Street is the leading provider of research on Real Estate Investment Trusts (REITs). Their public market research encompasses over 128 REIT/Real Estate operating companies across North America and Europe. Green Street also has several private market focused products under its Real Estate Analytics suite (started in 2013) that focuses on private real estate sector valuations and analytics across major U.S. markets and that also features databases on the mall, strip center, and New York office segments in the U.S.
Geography	Green Street's public market research coverage universe encompasses 82 companies across North America and 46 across Europe.
Frequency and Dates of Coverage	Green Street has been valuing REITs via a properietary relative pricing model since 1989 and generating Net Asset Value (NAV) models (which include data on nominal capitalization rates) since 1986 for the U.S. It started coverage of European listed companies in 2008. Updates of company valuations, pricing model metrics, and valuation indices are done monthly.
Underlying Data Source	Green Street develops company-level NAV models that are in turn used to drive its underlying REIT pricing model. . The model is based on the logic that REIT valuation can best be assessed by analyzing separately the two key components of value: 1) the market value of assets and liabilities; and 2) the present value of future investment opportunities. In addition, Green Street's model attempts to quantify the premium/discount that should be attributed at a company-level to franchise value, governance, executive compensation, and balance sheet management. These NAV values are also used to drive Green Street's Commercial Property Price Indices.
Completeness of Panel	Green Street's North American coverage universe represents USD 750 B of market capitalization, which is approximately 75% of the U.S. equity REIT universe market capitalization of USD 994 B as of September 30, 2016. The European coverage universe represents EUR $190 B of market capitalization, which is approximately 94% of the FTSE EPRA Developed Europe index equity market capitalization of EUR 205 B as of September 30, 2016.
Physical Property Characteristics	Green Street's NAV models contain cap rates in limited cases down to the property-level (i.e., malls, New York office), but most NAVs value REIT properties at a geographic level (i.e., Manahttan portfolio cap rates). In these cases, cap rates are estimated based on recent transaction comps, management commentary, third-party data providers, and broker contacts in the respective markets. For those sectors with property-level detail, the data is contained in three databases: 1) Green Street's Mall Database, which comprises information and insights into over 1,300 regional malls and other retail properties across the U.S., with a primary focus on REIT-owned properties; 2) Green Street Advisors' Strip Center Database, which offers information on 2,500+ REIT-owned strip center properties across the U.S.; and 3) Green Street's Manhattan Office Database, which contains data on 100+ properties (including some important privately held properties) representing 120+ million square feet and an estimated USD 40 billion in value.
Economic Performance Data	Equity market valuation measures provided by Green Street include implied capitalization rates, dividend/funds-from-operations/adjusted-funds-from-operations yields, and premia/discounts to NAV. Asset-level valuation metrics include nominal cap rates, unlevered IRRs by sector, and capital expediture models.
Ownership and Financial Structure Data	Green Street has detailed ownership information for all assets held by U.S. and European REITs to the extent that the information is publicly available. Data on privately held mall properties is available in its mall and strip center databases.
Caveats	Green Street commercial property price indices are based on estimated REIT asset values and changes in those. As such, they are neither transaction nor appraisal based enitrely, but rather a hybrid in the sense that Green Street's valuations are informed by transactions in the markets and thrid-party sources. Green Street private market data for specific geographies on cap rates is limited to markets where REITs own properties -- they do not track individual properties outside of the REIT-owned universe on a meaningful scale.
Research Use	Green Street does occasionally provide its historical time series data on NAVs to academic users but does not have a formal academic license

Academic Research Papers Using These Data	1. Clayton, Jim and MacKinnon, Greg, "Departures from NAV in REIT Pricing: The Private Real Estate Cycle, the Value of Liquidity and Investor Sentiment" (February 2002). Available at SSRN: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.178.876&rep=rep1&type=pdf 2. Driessen, Joost and Van Hemert, Otto, "Pricing of Commercial Real Estate Securities during the 2007-2009 Financial Crisis" (February 11, 2011). Available at SSRN: https://ssrn.com/abstract=1470249 or http://dx.doi.org/10.2139/ssrn.1470249 3. Clayton, Jim and MacKinnon, Greg H., "Explaining the Discount to NAV in REIT Pricing: Noise or Information?" (December 18, 2000). Available at SSRN: https://ssrn.com/abstract=258268 or http://dx.doi.org/10.2139/ssrn.258268 4. Cici, Gjergji and Corgel, John B. and Gibson, Scott, "Can Fund Managers Select Outperforming REITS? Examining Fund Holdings and Trades" (May 20, 2010). Real Estate Economics, Forthcoming. Available at SSRN: https://ssrn.com/abstract=1526771

Item	Response
Data Provider	**INREV (European Association for Investors in Non-Listed Real Estate Vehicles)**
External contact	info@inrev.org, T +31 (0)20 235 8600
Type of Service	INREV is Europe's leading platform for sharing knowledge on the non-listed real estate industry. INREV research reports include annual, academic and one-off reports to cover European and Asian non-listed real estate funds, capital raising activties of European non-listed real estate, analysis of funds of funds, sustainability performance, and fund termination strategies. The data are based on annual surveys from fund managers in North America, Europe, and Asia.
Short Description of Data	The dataset provided by INREV tracks fund level performance of properties of European NLREFs from 2001 on. Variables include (but are not limited to): fund characteristics such as size, yearly returns, leverage, age, vintage, investment style (core vs value-added), fund structure, and the INREV Index which measures annual Net Asset Value based performance. These variables may be measured on a quarterly and/or annual basis. Sectors are split into industrial, office, residential, retail, and others. See link below for the list of data fields:
Geography	About 80% of covered funds are from the U.K. (1), Netherlands (2), Germany(3), Italy and France(4).
Frequency and Dates of Coverage	Quarterly/annually since 2001 for all funds covered.
Underlying Data Source	Information in this database has been collected by means of questionnaires to debt funds managers and secondary sources such as company websites, press releases and industry news. The survey itself can be found at: https://www.dropbox.com/s/rp17c8hmlhex3k4/INREV_SDDS_3.0.xlsx?dl=0
Completeness of Panel	For largest countries listed above, there are over one hundred observations (fund-years) for which data for all explanatory variables are available.
Physical Property Characteristics	N/A
Economic Performance Data	At the fund level (e.g., annual return)
Ownership and Financial Structure Data	At the fund level (e.g., fund leverage)
Caveats	N/A
Research Use	INREV has sponsored research reports and has publicly stated its support for academic research (https://www.inrev.org/research).
Academic Research Papers Using These Data	Delfim, Jean-Christophe and Hoesli, Martin, Risk Factors of European Non-Listed Real Estate Fund Returns (May 30, 2016). Swiss Finance Institute Research Paper No. 16-37. Available at SSRN: https://ssrn.com/abstract=2793855 Dirk RSM, Hans 't, and Ville INREV (2007) Transparency in the European Non-listed Real Estate Funds Market. Journal of Real Estate Portfolio Management: 2007, Vol. 13, No. 2, pp. 107-117.
Link to field definitions	https://goo.gl/RHCwj7

Item	Response
Data Provider	**JLL**
Public contact	Reginald.ross@am.jll.com
Type of Service	Jones Lang LaSalle (JLL) has 180 corporate offices worldwide and operations in more than 750 locations in 60 countries. They have approximately 36,600 employees and provide comprehensive integrated real estate and investment management expertise on a local, regional and global level to owner, occupier and investor clients.
Short Description of Data	JLL regularly collects and tracks data on vacancy, capitalization rates, rental rates, absorption, and investement sale volume for over 50 markets nationwide. This data is collected mostly for commercial (office/retail/industrial properties, although inroads are being made into the residential space.
Geography	Nationals and regional. Particularly focused on majors US commercial markets
Frequency and Dates of Coverage	Quarterly since 1990 in most markets.
Underlying Data Source	Proprietary surveys from network of brokers/owners/managers along with public sources
Completeness of Panel	
Physical Property Characteristics	
Economic Performance Data	
Ownership and Financial Structure Data	
Caveats	Both the type of internal data collected and constraints on its usage are very similar across JLL and CBRE.
Research Use	Generally free for collaborative research use. See notes to CBRE.
Academic Research Papers Using These Data	One with NYU Pending

Item	Response
Data Provider	**NYC Mayor's Office of Sustainability (MOS)**
Public contact	MOS: (212) 676-3290; NYC Benchmarking Help Center: 212-566-5584, questions@benchmarkinghelpcenter.org
Type of Service	NYC Mayor's Office of Sustainability is a branch of NYC administration focused on developing cohesive sustainability plans for New York City. It's responsibilities include overseeing and implementing those policies with coordination with relevant federal, state and local agencies and communities as well as other partners such as non-profit organizations and academics. They provide research, analytics and data collection related to climate resiliency, energy supply, green buildings, greenhouse gas mitigation, transportation, and waste.
Short Description of Data	New York City energy benchmarking data is collected as part of Local Law 84 of 2009 (LL84), part of the Greener, Greater Buildings Plan (GGBP) that requires all privately-owned properties with individual buildings bigger than 50,000 square feet (sq ft) and properties with multiple buildings with a combined gross floor area more than 100,000 sq ft to annually measure and submit their aggregated annual energy and water use data to the City. It constitutes approx. 15,500 properties (23,000 buildings) in New York City which is around 2% of all properties, but 45% of total square footage.
Geography	New York City; although similar data structure/programs exist in Boston, Philadelphia, Chicago, Washington DC, San Francisco, among others
Frequency and Dates of Coverage	Annually, since 2010
Underlying Data Source	The data are generated through EPA Energy Star Portfolio Manager
Completeness of Panel	Depending on a year, it covers between 60 to 85% of properties required to report.
Physical Property Characteristics	Dataset contains such informtion as: ownership, address, size, use type(s), energy consumption, green-house-gas emissions, fuel type consumed, energy generated, building type specific characteristics, such as operating hours, occupant density, etc.
Economic Performance Data	N/A
Ownership and Financial Structure Data	Owner name; managing agent name
Caveats	Data are mapped to the NYC tax lot, what makes it ambiguous if there is multiple buildings on the same lot; data are self-reported
Research Use	Truncated version available through NYC Open Data
Academic Research Papers Using These Data	Kontokosta, C. E. (2015). "A Market-Specific Methodology for a Commercial Building Energy Performance Index," Journal of Real Estate Finance and Economics 51: 288-316. Kontokosta, C. E. (2013). Energy disclosure, market behavior, and the building data ecosystem. Annals of the New York Academy of Sciences, 1295(1), 34-43. The City of New York, New York City Local Law 84 Benchmarking Report September 2014. New York, NY, 2014. The City of New York, New York City's Energy and Water Use 2013 Report August 2016. New York, NY, 2016. The City of New York, New York City Local Law 84 Benchmarking Report September 2013. New York, NY, 2013. The City of New York, New York City Local Law 84 Benchmarking Report August 2012. New York, NY, 2012. C.E. Kontokosta, Predicting building energy efficiency using New York City benchmarking data, in: Proceedings of the 2012 ACEEE Summer Study on Energy Efficiency in Buildings American Council for an Energy-Efficient Economy, Washington, DC, 2012. Kontokosta, C. E., & Jain, R. K. (2015). Modeling the determinants of large-scale building water use: Implications for data-driven urban sustainability policy. Sustainable Cities and Society, 18, 44-55.

Item	Response
Data Provider	**MSCI (previously known as IPD)**
External contact	www.msci.com/real-estate
Type of Service	Tracks property and portfolio performance. Used for benchmarking, index construction, portfolio analytics for investors.
Short Description of Data	Data on income, expenses, capital value (appraised except when trasanction occurs) and returns on properties. Also, some lease data.
Geography	25 countries, along with various multi-national regions (e.g. Global, Nordic, CEE)
Frequency and Dates of Coverage	The following are the base dates and periodicity of indices produced by MSCI based on their data. However, underlying property level data may be available, with a smaller sample than full index, on a more frequent basis and/or longer time periods: Australia: 1984, quarterly Austria, 2003, annual Belgium, 2004, annual Canada, 1999, annual Czech Rep., 2004, annual Denmark, 1999, annual France, 1997, annual (semi-annual from 2007) Germany, 1995, annual Hungary, 2004, annual Ireland, 1994, quarterly Italy, 2002, annual (semi-annual from 2006) Japan, 2001, monthly S. Korea, 2005, annual Netherlands, 1994, annual (quarterly from 2007) New Zealand, 1999, annual Norway, , annual Poland, 2004, annual Portugal, 1999, annual South Africa, 1994, annual Spain, 2000, annual Sweden, 1983, annual Switzerland, 2001, annual UK, monthly, 1986, monthly (annual back to 2000) US, 1998, quarterly Smaller samples of data also exist for emerging markets such as China, Thailand etc. which are not large enough for a separate index but are aggregated into regional indices
Underlying Data Source	Data collected from investment managers and investors subscribing to MSCI portfolio analysis service.
Completeness of Panel	MSCI estimates their coverage as a percentage of properties held in "professionally managed portfolios" in each market they cover. As of 2015, coverage was: United States: 12.1% Canada: 37.2% South Africa: 56.7% Australia: 53.7% New Zealand: 52.3% Switzerland: 44.2% United Kingdom: 42.9% S. Korea: 36.4% France: 33.5% Finland: 31.9% Norway: 30.7% Netherlands: 30.6% Portugal: 30.5%

Physical Property Characteristics	Property type and sub-type, metro, address, geocode, MSCI also assigns unique identifying number. Also, sq. st., number of floors (for office and apartment), construction year, amenities.
Economic Performance Data	Net Operating income, capex, appraised value (transaction price when sold), income return, appreciation return, total return, vacancy. Lease information is available thorugh MSCI's IRIS product: data on weighted average remaining lease term, difference between in-place and market rents, probability of
Ownership and Financial Structure Data	Debt (amount and interest paid), joint venture details when applicable
Caveats	Returns are calculated using a monthly methodology (to be consistent globally, as the UK reports monthly returns) and then compounded to get the reported quarterly returns. Therefore the reported income and appreciation returns will not sum to the reported total return. For quarterly reported returns the difference will usually be small, but may be larger for markets where returns are reported annually.
Research Use	Due to confidentiality agreements with the managers and investors providing the underlying property data, researchers cannot access data at the individual property level. MSCI will accept requests for pulling specific data aggregates which they can then check to insure the supplied data does not contravene any required masking criteria. Any aggregated data point must involve at least five properties owned by three separate organizations before it can be released.
Academic Research Papers Using These Data	Data on the US has not yet been used extensively in academic research (the MSCI/IPD indices are relatively new to the market, and NCREIF has been the standard in the US for many years). At an index level, use of MSCI/IPD data is widespread and is the standard outside of the US - there is a voluminous literature using IPD indices for non-US research. For an example of property level data in academic research see "Risk Reduction and Diversification in UK Commercial Property Portfolios", Callender, Devaney, Sheahan, and Key, Journal of Property Research, 2007, Vol. 24, Iss. 4.
Other?	Investment strategy at purchase (e.g. development, redevelopment, reposition, lease-up, or stabilized).

Item	Response
Data Provider	**National Association of Insurance Commissioners (NAIC):** http://www.naic.org/prod_serv_financial_home.htm To download individual reports, https://eapps.naic.org/insData/
External contact	Phone: 816-783-8300 8:30am-5pm CT. Email: idp@naic.org
Type of Service	Statutory reporting requirement;
Short Description of Data	Schedule A contains information on each individual property owned, acquired, or disposed of including description of the property, location (city/state), book and market value (based on most recent appraisal), seller, actual cost, and date acquired. Schedule B contains individual mortgage loan, the vast majority of which are commercial. Each loan has a loan number, the location of the property, the interest rate, location, value of property, book value, and the year acquired.
Geography	Could be international properties or mortgages but only information for US subsidiaries
Frequency and Dates of Coverage	Rolling 10 year window (i.e., presently retain data back to 2006). Frequency is annual for holdings and quarterly for transactions (acquisitions and disposals)
Underlying Data Source	US insurance subsidiaries' statutory filings
Completeness of Panel	This is extremely comprehensive data of the holdings held by US insurance subsidiaries. Note that it is not comprehensive for the parent company - reporting is done at the subsidiary level.
Physical Property Characteristics	Mortgage loan numbers seem to be internal to insurer; no property IDs - each property identifier also appears internal to reporting insurer.
Economic Performance Data	Appraisals only give overall value and are not necessarily done every year; all transactions reported. Property-level data contain summary measure of income earned.
Ownership and Financial Structure Data	Know owner of the property if it is the insurance company or owner of the mortgage if it is the insurance company; for mortgages, can back out LTV
Caveats	See completeness. Forms, and thus reported fields, appeared to have changed though the years.
Research Use	PDF pages available at a price of $12.50 per company per year or $3.50 per company per quarter. CSV filings are available that could be used in Excel or Access. Unclear of the cost for a lot of filings. Might be able to merge on date acquired
Academic Research Papers Using These Data	None known.

Item	Response
Data Provider	**National Council of Real Estate Investment Fiduciaries**
Public contact	https://www.ncreif.org/staff.aspx
Type of Service	Produces and disseminates a suite of indices from property-level data. Members have access to aggregated property-level data through a web service and canned spreadsheeets. Commonly used for benchmarking and performance measurement. Several fund indices also available that can be linked to the property performance.
Short Description of Data	Detailed income, expense and capex data along with quarterly appraised values or transaction prices for mostly core tax-exempt institutionally owned CRE.
Geography	US
Frequency and Dates of Coverage	Quarterly since 1978
Underlying Data Source	"Data contributing" member institutions report quarterly data that comes from accounting or property management systems. Appraisal type, e.g., internal or external is identified as well as the sale price for sold properties. Property data for new members are backfilled, when available, in a "research" version of the database.
Completeness of Panel	Only member properties but all data contributing members required to submit all their properties including any non-core properties such as seniors housing or parking that are in the database but not in the official NCREIF Property index (NPI). Data on some value-added funds is now being collected to produce a fund level index where the underlying property data is not available. Based on CoStar data featured in Florance et al. (JREPM, 2010) and estimated in the textbook by Gertler et al., coverage of institutional CRE is about 10% (NCREIF had about $300B in property values in 2010). Based on Florance et al., office is over-represented (by about 3:2) and retail is under-represented (by about 2:3). Apartments and Industrials seem representative. Hospitality and speciality types are also under-represented. Geographically unbalanced representation (e.g., NYC CBSA under-represented and Dallas CBSA over-represented). The data reflects where institutional investors have allocated funds by property type and geographic area and not the overall distribution of commercial real estate. Being primarily core funds, the properties are heavily weighted to the top CBSAs in the nation.
Physical Property Characteristics	Unique "PROP" identifier is linked to property only for the duration of member's ownership (if purchased from another member, a new identifier is issued). Other available attributes: subtype, age, SqFt/NRA, #units, floors, zip code/msa. Address can be provided on a "need to know" basis to researchers. NCREIF working on adding geocodes for the properties.
Economic Performance Data	Property level variabels: vacancy, revenue (by source), expense (by source), capex (by source), purchase/sales, partial-sales, appraised values, cap rates. Derived variables are calculated (NOI, price changes, NOI growth, returns, etc.)
Ownership and Financial Structure Data	Owning institution/fund, fund type, manager, joint venture status, debt, debt payments, interest, ppl balance.
Caveats	Though generally reflecting industry best practices, the data contains significant incidence of outliers mainly due to the anomalies of accounting based data. Non-NPI data exhibits a greater rate of missing data.
Research Use	Academic membership and application with a research proposal is required to access property-level data. A non-disclosure agreement (NDA) must be signed by the researcher.
Academic Research Papers Using These Data	Research questions: Index construction and performance; property-level risk-return analysis. Sample academic papers: Geltner, D. and Goetzmann, W., 2000. Two decades of commercial property returns: A repeated-measures regression-based version of the NCREIF index. The journal of real estate finance and economics, 21(1), pp.5-21. Fisher, J.D. and Geltner, D., 2000. De-lagging the NCREIF index: transaction prices and reverse-engineering. *Real Estate Finance*, *17* (1), pp.7-22. Peng, L., 2015. The risk and return of commercial real estate: A property level analysis. Real Estate Economics. An, X., Deng, Y., Fisher, J.D. and Hu, M.R., 2015. Commercial real estate rental index: a dynamic panel data model estimation. Real Estate Economics. Pivo, G. and Fisher, J.D., 2011. The walkability premium in commercial real estate investments. Real Estate Economics, 39(2), pp.185-219. Technical documents: NCREIF DATA COLLECTION AND REPORTING PROCEDURES MANUAL Fisher and Dierkin (2009): Understanding and Querying the NCREIF Database

Item	Response
Data Provider	**Preqin**
External contact	www.preqin.com
Type of Service	Real Estate Online subscription; Private Equity Online (includes Cash Flow Data); Infrastructure Online subscription
Short Description of Data	Preqin provides data and intelligence on fund performance, fundraising, fund terms and conditions, fund managers and institutional investors in private real estate. Preqin purports to have the largest sample of fund level data (with performance metrics) amongst Burgiss and Cambridge. Preqin tracks 7,600+ funds across PE, CRE, and Infrastructure. For CRE specifically, detailed profiles for 3,400+ fund managers globally and 5,400+ unlisted real estate funds including limited partnerships, property unit trusts, LLCs, FCPs etc. and encompassing all strategies including core, core-plus, value added, opportunistic, debt and distressed and fund of funds are tracked
Geography	Preqin's distribution of "Fund Geographic Focus" (defined as the focus of the subject fund) is as follows: of the 5,100 real estate fund count in its universe as of 3Q15, 2,687 (53%) focused on the U.S., 1,451 (28%) focused on Europe; 566 (11%) focused on Asia; and the remaining 8% either focused on other regions or were unkown.
Frequency and Dates of Coverage	Preqin's data goes back to as early as 1984 in terms of reported vintages. Generally, there is a 3 month lag in GPs reporting performance information to Preqin after quarter end, and a 6 month lag for the LPs reporting to Prequin after quarter end, although there are some exceptions to this. Preqin bulk uploads information to its database once gathered from all sources on average once a month. At this stage, they regenerate the market benchmarks and assign funds with quartile rankings.
Underlying Data Source	Preqin obtains its data primarily from two sources: 1) Freedom of Information Act (FOIA) requests sent to public LPs (public pension funds and some endowments, in the US and UK); and 2) direct data from GPs. There is a 50/50 split between the two generally. As of 2Q16, Preqin has around 1,200 GPs globally providing them with data and around 500 UK and US FOIA compliant institutional investors. If a public LP has not invested in a fund and the GP declines to provide Preqin with performance data, data for that fund is not sourced and is therefore not available/tracked in its system.
Completeness of Panel	Preqin's focus is primarily on closed-end PERE funds. Although Preqin has profiles for non-closed end funds (open-ended, semi open-ended, etc), they don't focus on tracking performance of those fund structures. As such, of the 5,400+ unlisted real estate funds currently in its universe, Preqin has profiles for 4,150 closed PERE funds. Of these 4,150 funds, Preqin has performance data for 1,400 funds (33%) due to the fact that it obtains its information from public disclosures or select GPs that voluntarily elect to provide it. Weighted by fund size, Preqin has performance for around 60% of capital raised by the 4,150 closed-end PERE funds.
Physical Property Characteristics	Limited information on specific properties is available from fund quarterly report. However, Preqin added in 3Q16 a "Real Estate Deals" service module that tracks data for over 20,000 direct deals involving GPs. Over 15,000 underlying real estate assets are tracked andinclude features such property capitalization characteristics (i.e., leverage level; fund ownership) to the extent that information is identifiable.
Economic Performance Data	Measures include Net IRR, Called (%), Dist (%) DPI, Value (%) RVPI, Multiples. Preqin offers a "private equity cash flow" download (from the Preqin website) containing the "full cash flow information" on funds. As of July 2015, they report 859 buyout funds (602 in the U.S. and 257 elsewhere) and 680 venture funds (611 in the U.S. and 79 in the rest of the world). Preqin also provides fund benchmarking capabilities.
Ownership and Financial Structure Data	For funds with a profile in Preqin's database, information is available at fund level to the extent that it is disclosed as part of the public filings of the LP or is provided by the GP.
Caveats	Preqin's data is heavily reliant on Freedom of Information Requests (which are less frequently updated). There is also a large gap between Preqin's entire universe of CRE Funds (5000+) and those funds with performance data (1300), mostly due to limited data colleciton on non-closed end funds as well as on funds in which no or few public LPs are invested. Also, Preqin has less strict data requirements than other providers before entering a fund into their system, thus adding more funds into its population overall but possibly less accurate data
Research Use	Academic users have to subscribe as regular subscribers -- there isn't a separate academic license/policy

Academic Research Papers Using These Data	1. Krautz, Sebastian and Fuerst, Franz, Size Signals Success: Evidence from Real Estate Private Equity (July 15, 2015). Available at SSRN: https://ssrn.com/abstract=2649904 or http://dx.doi.org/10.2139/ssrn.2649904 2. Kiehela, Sami and Falkenbach, Heidi, The Performance of Non-Core Private Equity Real Estate Funds: A European View (September 2015). Available at http://www.bfjlaward.com/pdf/26030/062-72_Kiehel%C3%A4_RE_JPM_0919.pdf

Item	Response
Data Provider	**Real Capital Analytics (RCA)**
External contact	www.rcanalytics.com, info@rcanalytics.com
Type of Service	Transaction price tracking service, Typically used to track market trends.
Short Description of Data	Tracks all property transactions over $2.5 million in the US. Records buyer, seller, transaction price and some other data (see below for more details). Covers the following sectors: office, industrial, retail,
Geography	US (data can be broken down by region, state, city). RCA also covers transactions globally (146 countries outside the US) - the descriptions below refer only to the US data unless otherwise stated.
Frequency and Dates of Coverage	US data starts in 2001. (Coverage in UK and France (Paris only) also starts in 2001. Most other countries have coverage from 2007.) Indices are produced from the data monthly in US, but exact dates are provided for each transaction.
Underlying Data Source	Multiple cross-checked data sources are used: public records incl. tax documents, corporate records, and other; press releases; "scrapes"of web for information; CMBS market information on property; data obtained from brokers
Completeness of Panel	RCA believes that in the US they are at or close to 100% coverage for transactions over $2.5 million (for basic transaction data - coverage may vary for certain other types of data). Within the US: (1) RCA divides transaction in 4 sectors or main types: Apartment, Industrial, Office, and Retail. Apartment transactions represent 31% of the total, Industrial 20%, Office 25%, and Retail 24%. There are also 8 subtypes: Flex representing 6%, Garden 23%, Mall & Other 10%, Mid/High-rise 8% Office-CBD 6%, Office-Sub 19%, Strip 14%, and Warehouse 14%. (2) RCA classifies transactions in 117 markets from which the ones with more incidence are New York City (composed of Manhattan with 4% of the total and NYC Boroughs with 3%, making a total of 7%), Los Angeles (7%), Dallas (4%), Chicago (4%), and Atlanta (4%). Within these markets, Apartment represents 36% of total transactions (in these markets), Industrial 20%, Office 24%, and Retail 20%. (3) Investors are classified in 7 groups: CMBS, Equity Fund, Institutional, Private, Public, User/Other and Unknown, from which Private is the most important (over 60% transactions); and 25 types, from which Developer/Owner/Operator (over 50% of the transactions). Additional details are available at https://www.rcanalytics.com/Public/coverage.aspx.
Physical Property Characteristics	Database includes property address, GIS information, and unique property ID number. Also, each separate transaction is assigned a unique transaction ID number. Also includes property size, year built, year, if the property was renovated, and Walk Score. Property identifier "propertykey_id"; deal identifier "deal_id"; interaction between property and deal identifiers "property_id" (if a property is traded again, a new property_id is issued). There are several hundred thousand properties and deals in the US dataset for the entire available history (2001-). Other variables such as: type and subtype of property, address, country, city, county, zip code, year of built, # of sqft, latitude and longitude coordinates, etc.
Economic Performance Data	Transaction price, cap rate. Portfolio (if the property was traded as part of a portfolio transaction), occupancy rate (56% non-missing entries), price and adjusted price (if just a part of the building was traded), cap rate (27% non-missing entries). No data on cap rate components (source of cap rates not clear - perhaps parties to the transactions).
Ownership and Financial Structure Data	Buyer, seller, broker of transaction, indicator for whether property was bought/sold as part of a portfolio, lender for loans and refinancings, amount of loan (loan data is estimated by RCA to cover 60% to 75% of universe). Buyer and seller information includes category (e.g. pension fund, cross-border, private, institutional, etc.)
Caveats	Data is based only on properties that transact - transacting properties may have certain characteristics and not necessarily represent the broader universe of properties held for investment. The data may underestimate the extent of cross-border capital flows: Transactions record the listed buyers and sellers. A US-based fund manager buying a US property will be listed as a domestic buyer, even though the ultimate capital source for the fund could be non-US capital.
Research Use	RCA has a history of providng data for academic use. Access requires signing of non-disclosure agreement. Access can be arranged as either: (1) direct access to requested data, or (2) requests for RCA analysts to pull certain data sets.

Academic Research Papers Using These Data	Numerous examples including: 1. Demirci, Irem, Umit G. Gurun, and Erkan Yonder, "Real Estate Holdings of Public Firms and Collateral Discounts", Real Estate Research Institute working paper, 2016, www.reri.org 2. McAllister, Pat, and Anupam Nanda, "Does Foreign Investment Affect U.S. Office Real Estate Prices?", 2015, Journal of Portfolio Management, special real estate issue 3. Chichernea, Doina, Norm Miller, Mike Sklarz and Robert White, "A Cross Sectional Analysis of Cap Rates by MSA", 2008, Journal of Real Estate Research 4. Newell, Graeme, Alastair Adair, and Stanley McGreal, "Robustness of capital flows into the European commercial property markets during the global financial crisis", 2010, Journal of European Real Estate Research 5. Downs, D. H., & Xu, P. T. (2015). Commercial real estate, distress and financial resolution: Portfolio lending versus securitization. The Journal of Real Estate Finance and Economics, 51(2), 254-287.

Item	Response
Data Provider	**Reis**
Public contact	https://www.reis.com/
Type of Service	US CRE data provider specializing in creating submarket reports and comp reports (sales & rent) for its clients. Reports include history (up to five years) and forecasts.
Short Description of Data	Transactions, rents, and aggregated reports.
Geography	US
Frequency and Dates of Coverage	Quarterly data since 1980 (subscribers can go back to 1990, Reis' research reports produced internally can reach back to 1980)
Underlying Data Source	Fed through original sources rather than brokers, and privately collected by surveying owners and building managers to get property-level data (including performance). Many are clients, non-client sources are relationship based. Track transactions from public records and verify sales. New constructions from multi-source cross-referncing.
Completeness of Panel	Sales go back to 2001, performance to 1980. Now cover every transaction in US. Previously in 275 markets - increased through time (performance - 40 units and above; CA and AZ 20 units and above). All major property types. Recently added "special" property types (student housing, self storage, senior housing). Login to see the markets (and geographic coverage).
Physical Property Characteristics	Addresses, geocodes, subtype, age, SqFt/NRA, #units, floors, zip code/msa. Much more (e.g., renovations).
Economic Performance Data	Rental rates, vacancies, 5-year history available to clients. Going back further would require special access. Don't have cash flow from each building. Track operating expenses at property level. Don't track CapEx but CapEx is estimated (not clear how) when estimating cap rates. Last assessed value on sales transactions. Cap rates for sold properties - usually estimated (80% of the time).
Ownership and Financial Structure Data	Identity of "owner of record" but not type of institution. Transaction financing data. Details on financing terms vary. Mention made if part of a portfolio transaction. Tenant information (identities and sqft) - but no lease terms.
Caveats	
Research Use	Reis offers discounted (but limited) academic access to some institutions. Internally, the research group is mostly dedicated to serving specific client needs (only a few PhDs).
Academic Research Papers Using These Data	Goetzmann, William N., and Susan M. Wachter. "Clustering methods for real estate portfolios." Real Estate Economics 23.3 (1995): 271-310.

Item	Response
Data Provider	**SNL**
External contact	Sales@snl.com
Type of Service	REIT Analysis
Short Description of Data	Property-level data on all properties held by both traded and non-listed REITs; some non-REIT data
Geography	Global but information is primarily culled from SEC filings (10Qs, 8Ks, and 10Ks). Potentially biased towards US companies
Frequency and Dates of Coverage	1989-Present; Quarterly and Annual
Underlying Data Source	Public records, US is primarily from SEC filings
Completeness of Panel	Very complete for US REITs; unclear how complete for other
Physical Property Characteristics	Unique SNL property key; also has address including latitude and longitude
Economic Performance Data	Very detailed
Ownership and Financial Structure Data	Very detailed for parent company; has property-level lien information as well
Caveats	
Research Use	Would be very easy to merge with Compustat. High quality data on REITs
Academic Research Papers Using These Data	Riddiough and Steiner (2016) "One size does not fit all: REIT capital structure and firm value"

Item	Response
Data Provider	**StepStone Group**
Public contact	4275 Executive Square, Suite 500, La Jolla, CA 92037. (858) 558-9700
Type of Service	StepStone is a global private markets specialist overseeing over US$91 billion of private capital allocations, including approximately US$24 billion of assets under management. StepStone has internally developed a research database platform to track information that it receives from General Partners during due diligence processes. The platform also contains various research publications created by StepStone's 100+ investment professionals.
Short Description of Data	Quantitative and qualitative data at the GP, fund, and property-level, including strategy classifications, geographic focus, net and gross performance, meeting notes, fundraising data, etc. For funds that StepStone has conducted due diligence, StepStone's Assessment is available, which describes the merits and risks of the fund's organization, historical performance, strategy, and structure.
Geography	Global
Frequency and Dates of Coverage	Coverage dating back to the early 1970s. Updated on a quarterly basis or when funds are actively fundraising, depending on the data type
Underlying Data Source	Data sourced from General Partners and supplemented with other third party databases
Completeness of Panel	Over 4,600 Real Estate funds across over 1,900 General Partners
Physical Property Characteristics	Property-level performance is available
Economic Performance Data	Various measures of fund performance (e.g., net IRR, TVPI, DPI) and deal/asset performance (e.g., gross IRR, gross TVM).
Ownership and Financial Structure Data	Lists the historical GP investors for individual properties. Lists LPs and their commitments to individual funds.
Caveats	Pretty messy in spots, but has huge potential. Lots of gaps because it is basically a bunch of snapshots from due diligence process. So it has advantages and disadvantages.
Research Use	The website is accessible with login credentials provided by StepStone. Data on the web platform can be exported to Excel. Custom exports are available upon request.
Academic Research Papers Using These Data	Brown, G., Gredil, O., Kaplan, S., 2016. Do Private Equity Funds Manipulate Reported Returns. Working Paper.

Item	Response
Data Provider	**Trepp**
Public contact	See http://www.trepp.com/contact
Type of Service	Provides data, tools, and research support for CMBS deals, securities, and individual loans.
Short Description of Data	Current and historical CRE collateral information for CMBS loans.
Geography	US - entire CMBS market
Frequency and Dates of Coverage	Continuous. Roughly 20 years of data.
Underlying Data Source	Information is scraped from prospectuses and other public information about individual loan underwriting.
Completeness of Panel	Captures virtually the whole US CMBS and CLO loan universe.
Physical Property Characteristics	Address, size, age, rentable area, renovation year, units, type, condition.
Economic Performance Data	Pro forma valuation details needed to underwrite loan as well as appraisal values.
Ownership and Financial Structure Data	Owner, loan details + terms.
Caveats	
Research Use	Trepp often contracts with academic institutions to provide data and platforms for pedagogical and research purposes.
Academic Research Papers Using These Data	Lots of research employing this dataset. Some recent publications are listed below: An, Xudong, and Gary Pivo. "Sustainable Development and Commercial Real Estate Financing: Evidence from CMBS Loans." (2016). Akat, Muzaffer, et al. "Estimating Default and Defeasence Probabilities for Mortgage Securities." Real Estate Economics (2016). Ghent, Andra, and Rossen Valkanov. "Comparing securitized and balance sheet loans: size matters." Management Science (2015).

Item	Response
Data Provider	**Xceligent**
Public contact	https://www.xceligent.com/about-us/
Type of Service	Provides detailed property level data to virtually every kind of real estate professional that requires highly granular information. Proactive research on property characteristrics, property contacts (owners, managers, brokers), availability (lease or sale), tenant roll, historical sales + leasing activity.
Short Description of Data	Detailed physical, economic, and ownership characteristics for Office, Industrial, and Retail properties. CoStar competitor in covered markets. Some limited (market specific) availability of multifamily though coverage of mixed-use properties is good.
Geography	US - roughly 50+ MSAs. See website "coverage" link.
Frequency and Dates of Coverage	Data since 1999 at the earliest. Market-dependent. Sales comps might be back-filled (lease info if available). Update frequency based on listing/transaction report otherwise monthly.
Underlying Data Source	Physical and virtual site inspection, public records, Xceligent personnel research tenant rolls and occupied space. Listing info from network of brokers.
Completeness of Panel	Representitiveness within the 50 MSAs (by property type and class): everything over 5000 sqft and anything that has transacted. Generally, the larger the property the more uniform the detail. Especially for properties that have been "covered" by brokers.
Physical Property Characteristics	Incredible level of detail: E.g., parking spaces, # floors + sizes of smallest/largest floors + suite-level details, units, 3 major tenants, main construction material, HVAC system, identifier for building complex, gross land area net of pad, constructin details, elevator banks, traffic counts, amenities, zoning, tax assessment, anchors + their sqft, # restrooms, parcel identifiers, photos, geocodes.
Economic Performance Data	Sales + "verified sale" info: Transaction details, broker, sale date and signed contract date, tenants at sale, reasons for sale, down payments. Sale listing details: price and cash flow attributes. Details on leases: rates, escalations/contingencies/step schedules, tenant names/info, unit physical details, broker, concessions, quality, shared costs, listing details.
Ownership and Financial Structure Data	Ownership structure - Xceigent attempts to understand all oweners of records and their ownership shares, though this is still a "work in progress". Liens (details on amounts, rates and maturities). No information on debt terms.
Caveats	Revenues and expenses (including CapEx) info is generally only "at sale or listing" and only if broker includes info (including actual rents, TI or concessions). Generally, data is limited to items that the broker is willing to share. Consequently, calculating NOI would be a lot harder than, say, with Trepp.
Research Use	Everything in Xceligent is in principle available to paying customers.
Academic Research Papers Using These Data	Not aware of any

Appendix B. Current CREDA Working Group Members

Tom Arnold
Head of Americas-Real Estate
Abu Dhabi Investment Authority (ADIA)

Prof. Bob Connolly
Kenan-Flagler Business School
University of North Carolina at Chapel Hill

Prof. Jeff Fisher
Kelley School of Business
Indiana University

Prof. Andra Ghent
Wisconsin School of Business
University of Wisconsin - Madison

Preetesh Kantak
Kenan-Flagler Business School
University of North Carolina at Chapel Hill

Prof. David Ling
Warrington College of Business
University of Florida

Greg MacKinnon, Ph.D.
Director of Research
Pension Real Estate Association

Prof. Tomasz Piskorski
Columbia Business School
Columbia University

Prof. Jacob Sagi
Kenan-Flagler Business School
University of North Carolina at Chapel Hill

Prof. Stijn Van Nieuwerburgh
Stern School of Business
New York University

Prof. Greg Brown
Kenan-Flagler Business School
University of North Carolina at Chapel Hill

Andrea Marie Chegut, Ph.D.
Center for Real Estate
Massachusetts Institute of Technology

Professor David M. Geltner
Center for Real Estate
Massachusetts Institute of Technology

Prof. David Hartzell
Kenan-Flagler Business School
University of North Carolina at Chapel Hill

Prof. Constantine Kontokosta
Center for Urban Science and Progress
New York University

Prof. Crocker Liu
School of Hotel Administration
Cornell University

Prof. Joe Pagliari
Booth School of Business
The University of Chicago

Prof. Tim Riddiough
Wisconsin School of Business
University of Wisconsin – Madison

Prof. Sheridan Titman
McCombs School of Business
The University of Texas at Austin

Prof. Susan Wachter
Wharton School
University of Pennsylvania