

MILLENNIAL HOUSING COMMISSION

OUTLINE OF KEY ISSUES: INDUSTRY DATA WAREHOUSE

OVERVIEW

This paper summarizes key issues surrounding whether, and if so how, to create a multifamily data warehouse.

KEY ISSUES

1. **General Industry Benefit.** The prime rationale for the warehouse has always been to strengthen the primary and secondary multifamily mortgage markets, and the markets for multifamily equity capital, by improving transparency, facilitating benchmarking, and clearing the way for more standardization on better underwriting approaches. It is generally agreed that only the government can “seed” the creation of the warehouse. *All of this continues to be valid.*
2. **Barriers to Achieving the Benefit Now.** No one entity has sufficient data to provide an acceptably broad view of the multifamily stock and its financing. Moreover, those who do have large databases are reluctant to share their data due to concerns about confidentiality. Finally, the different databases are not in compatible formats and are not consistently quality-controlled. Hence the need for a respected, neutral entity to collect data, provide quality control, and disseminate information while protecting confidentiality of property-level data. *These barriers continue to be applicable.*
3. **Translating the General Benefit into Revenue.** The beneficiaries of this general transparency and benchmarking benefit are too diverse, and the benefit too indirect, to permit a fee-based revenue stream (in theory, the warehouse could be funded by a small fee assessed against all mortgage originations, but there is near-zero likelihood of that occurring in the real world). *Thus, in all likelihood, this general benefit would have to be funded either by the federal government or by the GSEs. As a practical matter, the sponsoring agency(ies) would have to make a credible long term commitment, else potential data providers might reasonably decide that the effort will fail and thus decide not to provide data.*
4. **Potential for Other Fee-Based Revenue.** There is some appetite in the industry for one or two simple products:
 - 4.1. **“Live” Market Data.** A report that summarizes recent revenue-related activity (rent increases, vacancy rates, turnover, ...) for, say, garden apartments in a particular price range in northeast Atlanta. *Generating this report requires frequent reporting of data by subscribers, and it is not clear that (a) the potential revenue justifies the incremental costs, or that (b) subscribers would, in fact, contribute data this frequently. Thus, this should not be viewed as a potential major revenue source, at least not initially. The warehouse design should permit*

this capability to be added later, however¹.

- 4.2. **Operating Expense Benchmarks.** A report that summarizes actual expenses (per unit, per square foot, percent of income) for properties that meet the subscriber's criteria (number of units, general location, resident profile, structure type, ...). *This is a viable product that would have modest but real value to subscribers. This product is superior to the more general income / expense studies now available because this product is customized to the user's specifications. At best, however, this is a modest volume / modest price product.*
- 4.3. **Limited Potential For Fees to Data Providers.** It is quite likely that data providers will expect to receive reports for free, in exchange for having provided data (which involves real cost to them). *Inasmuch as the data providers themselves are a large fraction of the potential market, it is unlikely that the warehouse will be able to support any significant fraction of its costs from user fees.*
5. **Conclusion re: Funding Approach.** *If the Commission recommends creation of the warehouse, the Commission should make clear that government and GSEs must be willing to cover the entire cost of the warehouse, on an ongoing basis. To launch the warehouse on any other basis would not be prudent.*
6. **Protecting Confidentiality of Raw Data.** The warehouse would, in any event, include minimum sample size constraints so that subscribers would have to specify a universe containing at least, say, twelve properties², so that the warehouse's report could not be traced to an individual property. The MFHI effort ran into a further concern on the part of the GSEs that their data be very strongly protected against unauthorized access. *It is reasonable to assume that similar levels of protection will be needed in any new approach, thereby adding to the cost of operation.*
7. **Competitive Concerns of Existing Private Data Providers.** A number of companies and trade associations publish data and/or provide customized research and reports on multifamily income and expenses. Examples include IREM, NAA, ULI, and M/PF Research. It will likely be essential to provide reasonable assurance that the warehouse effort is not intended to erode their competitive positions. There are two potential concerns.
 - 7.1. **Competing Products.** One is that the warehouse will offer products (with or without charge) that will compete with the existing products of private data providers. *The warehouse could avoid this by agreeing not to publish income and expense books.*
 - 7.2. **Obsolescence.** It is also possible that the existence of the warehouse will undermine the viability of the existing IREM / NAA / ULI income and expense studies; if IREM / NAA / ULI believe this to be the case, they might oppose the warehouse for parochial reasons. *The existing annual books might well become dinosaurs if the warehouse succeeds in developing high quality data, from a high*

¹ As a matter of database design, it is sufficient to design the initial database so that the later addition of quarterly or monthly data collection can be easily incorporated into the existing database.

² The correct threshold will be determined statistically, to give appropriate assurance that property-level data will not be identifiable.

volume database, in highly useable form.

- 7.3. **Potential Economic Imperative to Compete Later.** If the warehouse needs to be economically self-sustaining, private data providers will fear that the warehouse will eventually compete with the private data providers even if there are no immediate plans to do so. *This was one of the factors in the demise of MFHI. If the warehouse is government-funded / GSE-funded and focused on the primary and secondary mortgage markets, the private market-study providers might not feel threatened. Similarly, the warehouse could position itself as the provider-of-choice of raw data, thereby developing symbiosis with the private market-study providers.*
8. **Potential Intra-HUD Issues.** The Real Estate Assessment Center might be concerned that HUD support of the warehouse would tend to reduce the rationale for future funding to REAC. Similarly, it is possible that the Office of Housing and REAC might not be able to come to agreement on their respective roles vis-à-vis the warehouse. The MFHI experience showed the need for consistent, strong, and well-placed support within HUD.
9. **Conclusion re: Platform.** *The warehouse should have a charter that makes clear the public purpose nature of the warehouse and that satisfies stakeholder concerns. Placing the warehouse within a university or not for profit or governmental institution would be a useful but not absolutely essential step in this direction.*
10. **Scope of Data.** Categories of data that could be maintained in the warehouse include:
 - 10.1. **Annual Income and Expense, Most Recent Year.** Clearly, the warehouse will want to collect annual income and expense data.
 - 10.2. **Historical Income and Expenses.** Useful, but involving significant effort by data providers (to extract and provide) and warehouse staff (to convert, clean, and incorporate into the warehouse). *It is not clear that the benefit of historical data is sufficient to justify the added cost; instead, the warehouse will accumulate historical data over time.*
 - 10.3. **Property Characteristics.** Similarly, the warehouse will want to collect static data on property characteristics (unit mix, structure type, age, ...). The MFHI effort built in the capability to accept a very extensive range of data elements, which proved to be a problem both in terms of cost to develop the system, and in terms of acceptance among potential subscribers (some of whom were put off by the scope of data requested). *The aim should be to specify a minimal set of static data and loan data, designed for maximum benefit at minimum cost. The temptation to create a comprehensive set of property characteristics data should be resisted, at least initially.*
 - 10.3.1. **Location.** In the MFHI effort, Zip codes proved problematic because of the frequent changes in Zip codes for fast-growing areas. *Perhaps an alternative geo-coding approach, based on street address, would be better both because street address does not change, and because it is a finer level of geo-coding.*
 - 10.4. **Loan Data.** Data on loans (rate, term, origination date, amount, other financing terms) could be included as well, as this would be of considerable value from a secondary market perspective (helping to define the universe). Conversely, loan data are not vital for the other potential purposes of the warehouse. *This calls for*

a strategic decision about the objectives of the warehouse. It seems that if either HUD or the GSEs were to fund the warehouse, including loan data would be a high priority.

11. **How Data Are Accepted.** The MFHI experience teaches that data should be accepted in whatever form is most convenient to the subscriber. Warehouse staff are much better than subscriber staff at integrating data into the warehouse format, thus it is more efficient to have warehouse staff perform this function. Moreover, it is important to minimize the workload on subscribers so as to maximize participation. Experience also dictates that warehouse staff, rather than data providers, assume primary responsibility for quality control. *This implies additional warehouse staff, to take data in the form most convenient to subscribers and convert it into the form used in the warehouse.*
12. **How Data Are Provided For Research Purposes.** The warehouse will receive requests from researchers. Will the warehouse provide raw data? If so, under what confidentiality protections? *The latter question becomes particularly important given the legitimate GSE concerns about proprietary data.*
13. **Web-Based or Not?** The warehouse would certainly maintain a Web capability to receive data requests and would certainly supply benchmark data by email. Whether the warehouse itself should be Web-based is less certain. *It seems best not to Web-base initially, holding open the potential to do so later. In particular, Web-basing the warehouse would raise a host of data security issues; whether the warehouse could be made sufficiently hacker-proof is questionable.*
14. **Asset Matching.** Data for a particular property might be submitted by more than one subscriber (e.g. the owner, the managing agent, and the lender). A perfectly rigorous system for screening out multiple submissions is quite expensive. *It seems best to do only a simple screen based on location, property name and number of units.*
15. **Staffing.** The warehouse would need the following full time staff:
 - 15.1. **Chief Executive Officer.** Externally focused, on data providers, funding sources, and the capital markets. *Full time initially, not necessarily full time on an ongoing basis.*
 - 15.2. **Chief Operating Officer.** Internally focused. Manages staff. Manages contracts with software development consultants. Develops products.
 - 15.3. **Data Manager.** Working with data providers, cleaning data, resolving discrepancies, populating the warehouse, performing queries. *May need a second person during the initial building of the database.*
 - 15.4. **Administrator.** Receivables, payables, correspondence, general office support.
16. **Likely Annual Costs (after the first two years).** \$500K has been suggested as a potential annual cost. \$600K to \$700K annually seems more likely (once stabilized).
17. **Likely Start-Up Costs.** \$1.5 million has been suggested as a potential start-up cost, over and above the stabilized annual cost (spent over a one to two year period). This seems a reasonable estimate, provided that the warehouse itself does not reside on the Web.