Program

• Welcome
• Description of NYU’s Data Governance Initiative
• Comments from Representative Data Trustees
• Interactive Exercise
• Questions
Good afternoon. I’m Fred Cohen, and I’m the Associate Director of Data Governance and Institutional Research. In this position, I also carry out the role of Chief Administrative Data Management Officer for NYU.

Today, I’m going to focus on three main topics:
1. What is Data Governance, and why do we care?
2. Data Governance structure and roles at NYU
3. and What’s Next

So let’s begin with What is Data Governance, and why do we care?
When we talk about a new term, often times we want to begin with a definition. Here is a definition of data governance – which is a little longer than I might like, but which I think conveys the full breadth of what the concept encompasses.

Data Governance is a set of processes that ensures that important data assets are formally managed throughout the enterprise.

Data Governance ensures that data can be trusted and that people can be made accountable for any adverse event that happens because of low quality data.

(continued)
It is about putting people in charge of fixing and preventing issues with data so that the enterprise can become more efficient.

Data governance also describes an evolutionary process for a company, altering the company’s way of thinking and setting up the processes to handle information so that it may be utilized by the entire organization.

- Steve Sarsfield, The Data Governance Imperative

It is about **putting people in charge of fixing and preventing issues with data** so that the enterprise can become more efficient.

Data governance also describes an **evolutionary process** for a company, **altering the company’s way of thinking** and **setting up the processes to handle information so that it may be utilized by the entire organization**.

Some of you may be thinking that based on this definition, you are already doing tasks which seem to fit under data governance. That’s true. And that’s really something you should keep in mind... we are not necessarily talking about adding to your work, but rather recognizing, and giving a name to, the important work already being done.

Some clarifications:

- We are talking about ADMINISTRATIVE data – not research data, not academic data (Learning Management Systems). There are other groups around the University focusing on those areas.

- And at this point in time, we are working solely with the University-wide administrative offices. We know that schools have administrative data as well – both extracted from central systems as well as developed on their own – and we will include representatives from those areas in the future.
Where Are We Beginning From?

- Development of UDW+ has driven early data governance efforts
  - University Data Management Policy
  - Identification of Data Stewards for UDW+ related issues
  - Decision Support Group

So – where are we beginning from:

In order to properly develop and implement the UDW+, it was necessary to begin data governance efforts. These included:
- the development of the University Data Management Policy
- identification and training of Data Stewards for business areas as their data moves into the UDW+, and
- the development of the Decision Support Group to assist the Data Stewards in resolving data issues and to provide reporting and analytical assistance to UDW+ users

This is a good point to acknowledge the efforts of the PSO and DSG which have begun to move NYU to a more structured and advanced state of data governance. We are now going to leverage that work and expand upon it.

Let’s talk a little bit about what goes into a Data Governance Initiative
Data governance is built on four pillars:

The first pillar is **policy**. You received a copy of the University policy on Administrative Data Management with your invitation to this kick-off session. Part of the future effort is to create area-specific policies which implement data governance for each office, in line with the over-arching policy.
There are existing data-related processes in place, and as part of the Data Governance initiative, there may be changes to existing processes or new processes developed. For example, how do we identify and fix erroneous data – both in the transactional source systems and in the reporting systems? What is the process for adding new values to existing fields, and for conveying that information to all concerned... or even involving those who will be affected before the new values are added.
The third pillar is **Tools and Technology** which are used to assist with the tasks related to data governance and data quality. For example, as part of the data load process to the University Data Warehouse, there is the checking of data as it is loaded, and reports of erroneous data are generated. Technology can also be used to create better documentation of the meaning of data, to help ensure that we are all on the same page when referring to particular terms.
The final pillar is **Accountability**: Who is responsible for clean, complete data? It is all of us, in various capacities – from formulating policy to working with and cleaning data. Ultimately, data stewardship is everyone’s responsibility.

Let me emphasize something here: Data Governance, including data quality, despite what many people think, is not solely an IT responsibility. In fact, most of the onus falls on the business side of the University – the ones who create and use the data. IT can support the business in providing means to prevent, catch and correct bad data. But IT needs to know from the business how to define what is “good” or “bad” data.

Let’s move on to discuss why we are beginning this Data Governance initiative.
First, data governance will help to ensure the best data is available for making the best decisions possible for the University.

And one of the ways to ensure better quality data is to have more complete data.
More Complete Data:

- Tracking of faculty with multiple appointments
- Incomplete academic record attributes when using a “quick admit process”
- All tracking dates entered into sponsored research system so can establish benchmarks

We know that there are cases where missing data impedes the ability to report accurate information –

- More and more faculty members have interdisciplinary appointments, and are affiliated with more than one department. Each department wants to track the faculty as being part of their department, which was not possible in the legacy HR system and will hopefully be available in PeopleSync.

– When the Campus Solutions Student Information System was first implemented, a “quick admit” process was used to deal expeditiously with problems from the conversion. However, this process created its own problems when not all information was entered into a student record.

– Not all date fields are completed in the Sponsored Research system, making it difficult to know when a proposal was submitted.

One of the goals of the data governance initiative will be to discover and document where incomplete data enters our information systems, and create checks and validations to prevent it from happening in the future.
Why a “data governance initiative”? 

• Better data for better decisions  
  – More complete data  
  – More accurate data 

We also want to ensure more accurate data.
This is data from our student information system, and shows the format of data in the Campus ID field. We all have a campus ID, and know that it is a 9 character field, starting with “N”.

The first row indicates that there are 2,649 rows of data with no Campus ID. That’s an immediate problem which should be looked into.

The next row shows that over 4 million rows – 99.94% - are in the correct format of one letter followed by eight digits. We still don’t know if those are all correct – some may start with an “A”, or just be wrong numbers, but at least they are in the proper format.

The last five rows show that there are 39 cases of bad data – no question about it, as these values are not formatted properly.

Again, with a data governance structure in place we will be able to both identify data fields where the data may be inaccurate, and develop processes to prevent inaccurate data from even entering our administrative systems.
Why a “data governance initiative”? 

- Better data for better decisions
  - More complete data
  - More accurate data
  - Common understanding

And, we want to be sure we have a common understanding of terms and data usage.
# Common Understanding

## How many students are there at NYU?
- Why would a school dean and the University Registrar report different counts of students to the Provost?

## Why did the Mars Climate Observer fail?
- What caused the Mars Climate Observer to fly too low into the Mars atmosphere and be destroyed?

---

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>NYU</strong></td>
<td>Institutional Research And Data Integrity</td>
</tr>
</tbody>
</table>

---

- ok – why would a dean and the Registrar report different numbers of students?
  
  a. Only undergraduate, or only graduate, or combined undergraduate and graduate
  b. If both are reporting only undergraduate –
     i. may include those on Study Abroad or not
  c. If both are reporting only graduate
     i. May include those maintaining matriculation or not ]

- does anyone here remember the Mars Climate Observer failure from 1998?

The space craft’s mission, to study the Martian climate, came to an abrupt end when the lander for the Mars Climate Orbiter spacecraft flew too low in the Martian atmosphere and was destroyed. What caused this navigational failure? A NASA subcontractor used imperial units (as used in the United States), rather than the NASA-specified metric units (as used in Europe) to calculate the lander’s altitude.

These are two examples of situations where in each case the participants thought they all knew what they were talking about – but in the first we see that there are many definitions for “students” and in the second we find that a number without an agreed upon understanding of units can lead to disaster.

One of the things we will be striving for as part of the data governance initiative is to make sure that all assumptions are made explicit and that all definitions are clear. This will be accomplished by the data stewards participating in documenting how data fields in our systems are used, and creating a business glossary.
Why a “data governance initiative”? 

• Better data for better decisions
  – More complete data
  – More accurate data
  – Common understanding
• With better data, transactions can be processed more quickly, with no need for re-work or seeking the correct information
• Faster technical development time

Having better data also means that the information is available more quickly, and that it is correct – meaning that it is handled more efficiently. When multiple records are found in the Student Information System for the same person, it is a time-consuming and difficult task to merge the records, involving de-enrolling all courses from the incorrect ID and enrolling the person in the same courses under the correct ID, entering grades, ensuring that billing is not affected. Each case where one individual has more than one record, creates work for numerous offices to correct. The data governance initiative will work with the data stewards, data domain stewards and data trustees to develop processes to reduce or eliminate situations where these duplicate records are created.

And with better documentation of data elements and business processes developed as part of the data governance initiative, our information technology colleagues will be able to develop new applications more quickly, since the information they need regarding where to find particular data, or the contents of fields in the data warehouse, will be documented.
What are some of the expected outcomes from this effort?

- data policies which will provide guidance on how data should be handled, who should have access, and how bad information should be found and corrected.

- data standards to establish benchmarks for the quality of data – what is the proper format of data fields, how should data definitions be constructed, and how should data fields be named to be clear to everyone what the field contains.

- we have “resolved data issues” now. But we would like to begin to track those issues so that we can determine if there are patterns of bad data. If we can identify patterns, hopefully we can create means to prevent bad data from ever entering our information systems.

- as you have already heard, one key goal is to have higher quality information in our information systems so that better quality decisions can be made.

- and again as we’ve already highlighted, by having clearly defined data fields, there will be no ambiguity in what a particular field or value means, again improving data quality.

- Identifying the “official” source of data, so that if there are conflicts we can easily identify the authoritative and correct source.

- And with this improved data quality, the ability to meet management’s needs for information for decision making, beyond simply having operational data in transactional systems.

It’s an ambitious set of outcomes, and it won’t all happen at once. But with management’s support, we can make progress in making NYU more cognizant of the value and importance of clean, accurate data.
Agenda

• What is Data Governance? And Why Do We Care?
• Data Governance Structure and Roles
• What’s Next?
As you’ve read in the Administrative Data Management Policy, there’s a three tiered data governance structure in place:
The Data Trustees are at the top, with the Data Domain Trustees in the middle, and the Data Stewards supporting both of those groups. Additional players in the Data Governance initiative include the Data Custodians and myself, the Chief Administrative Data Management Officer.

There are other groups, not pictured here, which support data users. One example is the Decision Support Group which works with users of the UDW+ to resolve data questions for that particular environment.
At the top of the pyramid are the Data Trustees, who

- establish policies and procedures

- assign management accountability for the data within the administrative data systems,

- and who are the final arbiters for any issues rising through the data governance structure.
In the middle of the structure are the Data Domain Trustees, who

- are senior managers responsible for the transactional systems such as fame, PeopleSync, and Advance
- implement the policy established by the Data Trustees
- are responsible for assigning the Data Stewards, and
- serve as the first level of escalation for any data issues which cannot be resolved by the data stewards.
Underpinning the structure are the Data Stewards, who are:

- generally the operational managers responsible for the business processes and business rules associated with our administrative transactional systems, and

- are appointed by the Data Domain Trustees.

More specifically, those individuals in the role of Data Steward will be expected to -

Identify business rules for validating data  
Review data which fails the validation rules  
Participate in the development of data definitions  
Resolve data problems  
Administer security roles/approve access  
Communicate changes to business process or code sets which may affect others
To this point, we have really been focusing on the “business” side of data, and haven’t spoken about our technical colleagues – in ITS or in your own offices and departments. And this has not been an omission. Although we identified “Technology” as one of the four pillars of Data Governance, we really view Data Governance as the responsibility of the business side of the University. But we will be depending on the data custodians to support our data governance efforts.

The Data Custodians are

- the information technology experts who have as part of their roles

- overseeing the administrative transactional and reporting systems for the University.

In practicality, they will also be assisting the data governance effort by coding checks that data meets the appropriate business rules, as well as modifying source systems, if necessary, to prevent bad data from even entering transactional systems.
That statement – coordinates all activities related to NYU’s administrative data governance – seems so simple, and could actually be restated simply as facilitating the flow of information about data. Some examples of specific tasks include:

- Working with you to identify incorrect or incomplete data in our systems, to correct that data, and along with the Data Custodians to find mechanisms to prevent incorrect or incomplete data from initially entering our systems.

- Reporting to the Data Trustees on the state of the data in University administrative data systems

- Developing Master Data Management and Reference Data processes, including the maintenance of the Reporting Department Crosswalk, a tool being used to enable easier cross-system reporting.

- Developing standards for data definitions, and working with you to develop those definitions – both for individual areas and for cross-department terms.

- Working with you to establish data access policies for each functional area, and being a resource when non-standard requests are received.

- Publicizing the data governance efforts to the larger NYU community.
Agenda

• What is Data Governance? And Why Do We Care?
• Data Governance Structure and Roles
• What’s Next?
What’s Next?

- This is just the beginning
- Goal is better (more accurate) data in all systems
- Will continue to meet on a regular basis
- May have smaller working group meetings
- Will have training sessions

This is just the beginning
   - We wanted to let you know what is happening, and what the future holds

Goal is better (more accurate) data in all systems
   - Allowing the university to conduct its business more efficiently and effectively

Will continue to meet on a regular basis
   - Probably quarterly

May have smaller working group meetings
   - Especially when dealing with data definitions for common elements, or when business processes affect multiple offices

Will have training sessions
   - On various topics, including data profiling to find errors; data security requirements; writing good data definitions
Program

• Welcome
• Description of NYU’s Data Governance Initiative
• Comments from Representative Data Trustees
• Interactive Exercise
• Questions

Tony Jiga
Randy Deike
Stephanie Pianka
Andy Gordon
Data is both the “input” and the “output” of NYU’s financial operations, we use and transform data every day to both inform and service the NYU community.

from Stephanie Pianka
When data is inaccurate or processed incorrectly, the consequences can range from employee and client dissatisfaction to inaccurate financial statement and reporting with many potential negative consequences to NYU:

1. **Time & attendance** => Inaccurate Paycheck, Potential Tax Consequences, Dissatisfied Employees / Students
2. **Tuition & Fees** => Incorrect Bill, Dissatisfied Students, Parents, More Work for Bursar
3. **Vendor invoice** => Payment Errors, Shipment Delays, Re-work for AP, Procurement, SPA, Fiscal Officers
4. **New supplier** => Insurance risks, year end 1099 issues, vendor payment errors, payment delays
5. **Grant Awards** => Federal compliance requirements, fund drawdown timeliness, PI community dissatisfaction
6. **Space usage** => Tax-exempt bond proceeds compliance, accurate annual reporting on form 990 schedule K

Your role as a data steward is critical to ensure the integrity of our financial reporting and to maintain a high level of service to the NYU community.
UHR is responsible for the accuracy, privacy, and security of data relating to NYU employees and adhering to internal processes and policies and government regulations. Examples include:

<table>
<thead>
<tr>
<th>UHR</th>
<th>Data</th>
<th>Policies / Regulations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Applications</td>
<td>employee self service, contact information</td>
<td>NYU Data Management</td>
</tr>
<tr>
<td>Benefits</td>
<td>medical (incl. family) records, disability, leave, beneficiary</td>
<td>HIPAA, FMLA, ERISA, CORRA</td>
</tr>
<tr>
<td>Compensation</td>
<td>salary, incentive compensation, performance</td>
<td>FLSA, Form 195</td>
</tr>
<tr>
<td>Employee Relations</td>
<td>grievances</td>
<td>union contracts</td>
</tr>
<tr>
<td>Talent Management</td>
<td>applicant tracking, background checks</td>
<td>I-9, Affirmative Action</td>
</tr>
</tbody>
</table>

from Andy Gordon
- Quarterly Review of access privileges granted to HR Administrative Systems
- Implementation of Workday
- Internal Audit / Legal review
- Annual ERISA review of Health and Welfare and Retirement Plans
- HIPAA

Assess

Accuracy ➔ Access (Security) ➔ Interpretation (Integrity / Commingled Data)

If in doubt, give a shout!
Program

- Welcome
- Description of NYU’s Data Governance Initiative
- Comments from Representative Data Trustees
- Interactive Exercise
- Questions
OK – now comes the time for you to think about what you’ve just heard, about the problems you face in your office in dealing with administrative data, and in potential solutions.

At your tables we’d like you to have one volunteer to be a facilitator, to be sure everyone is heard from, and a second volunteer to be a scribe and reporter. First, we’d like each of you to take a few minutes and on the white paper note what data-related problems you currently face, and some ideas about how those problems might be corrected or prevented. After everyone’s had a chance to think about their own responses, everyone at your table should share your findings, and together determine if there are any similarities to the problems you face or the resolutions you envision. We’d like the scribe to jot down these commonalties on the blue paper.

After 10 minutes we’ll come back together, and have the scribe report the findings from their table. Depending on the time available, we’ll get to as many tables as we can. But we’ll be collecting the both the white and blue papers at the end, so that we can read all of the ideas generated.

Let’s get started. Again, begin by thinking about your own situation, and then come together to find similarities.
There’s been a lot of information packed into one session. And it’s the end of the day and you all would probably like to go home. But let’s hear if there are any questions based on what has been discussed today.
Thank You!