NORTHWESTERN
UNIVERSITY

A New Vision for Research Administrative Systems

Research Administrative Systems Working Group Report

Tom Board
David Keown
Warren Kibbe
Andrew Ludington
Steven Moyano
Luna Rajbhandari
Jason Schober

November 7, 2013

Contents

Executive Summary

Research at Northwestern is a large and growing endeavor: the University has more than $500 million in sponsored research grants, and the new medical research facility in Chicago is targeted to help attract an additional $150 million a year in new medical research funding. To support this endeavor, the University has an extensive inventory of administrative systems and a large number of specialized administrators, yet a common refrain on the research administrative process is that it is too unorganized, too unwieldy, and fails to address key business needs.

The current IT ecosystem for research administration is built around an inventory of multiple commercial off-the-shelf (COTS) transactional systems. These systems have grown greatly in number and functionality over the years, added great value, and will continue to play a critical role at Northwestern. However, these systems are "fully-featured", "niche" solutions, designed to run as free-standing individual systems, i.e. as islands isolated from one another. Within their targeted niche, these vendors have been successful in taking a functional area, integrating and unifying data, processes, and functionality *within* it, and growing their niche over time. In order for us to improve the overall research endeavor at Northwestern, we need to provide a framework that enables unification and integration across the enterprise's systems and functional areas.

Not only are the systems themselves "niche" focused, the overall IT infrastructure – within both the University and research administration – has been "niche" focused, prioritizing local Line of Business needs and resources over enterprise integration. The result is an infrastructure and a large set of services that provide many services but is fragmented and unwieldy, and becomes even more so as research funding, attendant regulations and requirements, and collaboration grow.

An approach that is premised on an enterprise-wide perspective and infrastructure is offered as an alternative: a set of systems that consciously functions as a portfolio of systems; is explicitly constructed to be aware of, and interact with, one another and the surrounding University infrastructure; is designed to be flexible to accommodate change on multiple levels; provides "unifying" tools/windows for both researchers and administrators; and is purposefully guided by an enterprise perspective and governance framework.

In order to accomplish this, work needs to be done in four main areas:

1. Our inventory of transactional systems needs to become a portfolio of systems, chosen initially, or modified if already in place, to work with each other and with our enterprise infrastructure.
2. Our enterprise infrastructure needs to provide a robust Services Architecture that enables data to flow in real-time and processes to move online as self-service functionality
3. We need to provide a set of core "unifying applications" that will sit on top of the transactional systems, and utilize the enterprise's enabling infrastructure to abstract the many transactional systems and processes into more easily manageable, personalized content and tasks.
    a. Data integration is needed in order to enable systems to "talk" to one another, to allow cross-functional area analytics, to enhance compliance auditing, and to reduce wasted effort in data entry and reconciliation.

b. A more feature rich and flexible Identity and Access Management infrastructure (which unifies a person's identity, minimizes disparate system log in requirements, and utilizes authoritative identities in federated systems to reduce identity management overhead and delays) is needed to allow cross-system and cross-institution interaction, and to facilitate compliance auditing.

c. Unifying views into this data and the services built on top of them are needed in order to bring much higher levels of efficiency to the business processes of research administration, and to drive improved decision-making via access to relevant data at appropriate times. Included in these applications are the integration of a research portal area within the University's portal and enabling the creation of better reports, dashboards, and analytic capabilities built over integrated data warehouse architecture.

4. Our prioritization and resource allocation processes need to become less Line of Business focused and more enterprise-focused, these processes need to be integrated within the existing IT Governance framework, and they need to happen at both the System Portfolio level and the data level.

These tasks will require great effort and coordination, and will compete for attention and resources with the Line of Business needs, which have been historically dominant. It is the consensus of the working group, however, that continuation of the current approach to research administration will ultimately be unsustainable and self-defeating within expected resources as the scope and pressures on research administration grow.

Included in the report and its appendices are a number of tools and suggestions to aid in movement towards these goals, including Next Step Recommendations (for some shorter-term pragmatic improvements, for more overarching architectural changes, and for the eIRB upgrade project), System RFP tools (a sample features matrix and Architectural Requirements that should be included), and data diagrams and flows.

This analysis and report could have been done for any functional area within the University, and the results would have been similar. Our hope is that research administration can be a leading area in a drive for institutional change. While research administration contains some of the biggest challenges in this discussion due to the inclusion of the FSM-NMFF-NMH relationships, it also has the advantage of being a relatively focused line of business, and hopefully has opportunities for making advances that might be more difficult to realize in more diverse parts of the University.

The Context

The administration of extramural research is a complex and wide-ranging set of processes that, taken together, comprise a comprehensive line of business for R1 universities. In the early days of government funding, the process of research administration was largely paper-based, with research administration generalists shepherding projects through to compliant completion. Over time, the scope of funding possibilities, the amount of money and resources involved in the research, and the requirements attached to administering the research grew so large that generalists gave way to specialists, and paper gave way to online systems.

The earliest of these systems were only individual applications built by local staff to address isolated problems/needs. Gradually, third-party vendors emerged to provide commercial off-the-shelf (COTS) solutions that provided more comprehensive packages of functionality, although these "niche" products are often still limited to a single area of responsibility (e.g. Pre-award grants management.) Each new level in this evolutionary trend has provided faster, more integrated, and more powerful approaches to managing the evolving needs and opportunities of research administration.

The evolution of research administrative systems at Northwestern has followed this trajectory to the point where we now have an extensive inventory of systems – both packaged and custom developed – that meet a wide array of the diverse and highly specialized needs of our research enterprise. Appendix A: Overview of Research Administrative Processes, Systems, and Data Flows, page 26, contains a list of the main administrative systems that enable these processes, a matrix showing the flow of data families across these systems, and multiple flow diagrams that show the complexity of the research administrative process (e.g. the life cycle of a Clinical Study).

What is readily apparent from even a cursory review of these diagrams is the complexity of the administrative processes, and of the data and systems that enable and support them. What is <u>not</u> readily apparent is that while this inventory of systems is far-reaching and powerful, these COTS systems are inwardly-focused, self-contained systems that provide a suite of functionality for a particular niche within the research administration process. No matter how "fully-featured" these systems might become, any collection of these isolated independent systems will leave (or even create) significant disconnects that constrain (or even undermine) the research process in a time when research administration needs to be more interconnected and flexible in order to meet the challenges of increasing compliance and privacy requirements, increasing collaboration in research (both within the University and with colleagues/research units at partner and peer institutions), increasing competition for funding, and the University's strategic goals for institutional expansion and sponsored research funding being at all-time highs.

The combination of this complexity and the new pressures on research administration begs for movement to another evolutionary level. COTS systems will, and should continue to be, a key piece of our IT strategy. However, if the goal is to get to a new level of research administration, business process conveniences need to be created through the integration of information across systems, and the ability to put data together as needed across the research span of business needs to be facilitated.

In order to do this, we need to leave behind the view that a grant can be independently administered across multiple systems – each with its own management priorities – and instead adopt the approach that a grant moves within an enterprise workflow of orchestrated functions. We will need to focus less

within a system, and more between systems, and this approach will require new forums and new benchmarks for making decisions about needs, priorities, and solutions from an enterprise perspective.

One benefit of the work and perspective that is being proposed here is that while the process examples illustrated in Appendix A are some of the more complex types of research that is done at Northwestern, the work that will be done to aid these processes can be leveraged across the institution, improving the process for less complex types of research right along with the more complex.

The Current State of Research Administrative Systems at Northwestern

Northwestern's current inventory of research administration systems (See Appendix A: Overview of Research Administrative Processes, Systems, and Data Flows, page 26.) revolves around a set of core systems, supplemented by a set of ancillary systems and undoubtedly by multiple uncounted shadow "systems" on spreadsheets, personal databases, and paper.

A top level profile of this inventory would include these characteristics:

1. Much of the research administration life cycle has online systems that support it – human and animal subject oversight, pre- and post-award financial management, clinical trial management, COI disclosure, research safety and compliance, space management, and effort tracking.
2. The systems themselves are "owned" by a variety of research business units who make functionality/investment decisions based on their own individual needs. There is no established cross-system mechanism for determining enterprise priorities that includes direct business users.
3. The state of the current systems obtained from third-party vendors varies widely – some are on the latest release of the vendor's software, some are multiple releases behind and difficult to upgrade.
4. There is no concept of Master Data Elements – an authoritative version of a data element that can be created/updated/deleted in only one place - across the set of systems. Today, the location of the most accurate version of a data element - e.g. people actually working on a project - can change as research moves through its life cycle, yet none of this is planned and little of it is documented.
5. When systems are "integrated", the solutions are "one-off" solutions that are asynchronous, hard to support, and non-scalable. The most widespread of these solutions are the batch data extracts that are used to share data between systems: data is extracted on some schedule, and then sent to the receiving system for subsequent batch importing. Each extract is a custom routine that is uniquely developed, is a relatively brittle solution to maintain on a day-to-day basis, adds time and cost to each software upgrade, and introduces time-lags into data utilization.
6. When data is not shared but is needed in multiple systems, it is re-entered, and subsequent changes to that data are not always transferred back to the authoritative system(s).
7. Permission to use research systems is managed and granted by different organizations using different identifying credentials. As a result, researchers must use multiple identities and

passwords to engage tools from different portions of the research support structure. This complicates the administration of grants by the researcher, duplicates effort by system administrators, and makes it very difficult to audit access records, implement single sign-on for portal deployment, or control access through policies.

8. While online workflows exist within a number of our research systems, there are few, if any, cases in which automated online workflows span systems. Likewise, data edited in one application rarely flows to other systems, or triggers updates to be undertaken downstream, and when it does, it is certainly not in real time.
9. There is no over-arching view or "portal" into this array of systems for people involved in the research process. Each person must learn (with widely varying degrees of help around them) the systems they need to use, how to get to each one, and what the next step in the process is.
10. Bringing data together for analytics is difficult due to the lack of shared data definitions, the lack of common institutional informational access policies, multiple "authoritative" sources of data, and data residing in different individual data marts that are "owned" by different business units, managed by different teams, and governed by different access control rules. These limitations make it difficult to enable self-service access to integrated reporting solutions, and the "work around" is for people to go to different reports in the BI tool or pull data directly from transactional systems, and then compile the information manually in a desktop application such as Excel or a shadow system.

An Alternative Vision

We are proposing a model that has a wider, enterprise-centric perspective premised on the following three primary assumptions:

- By integrating and unifying data, processes, and functionality in specific areas, the growth of these online systems over the last decade has greatly improved the University's ability to facilitate and manage the research process. There is every reason to believe that these "niche systems" will continue to grow in scope and number, and utilizing them should be a key foundational element for our portfolio of services.
- While key vendors have been successful and have emerged as leaders in their respective market sector(s), there will never be one vendor providing one overarching research administrative system. Therefore, the University's research endeavor will always be supported by an ecosystem of applications and vendors.
- While COTS systems will continue to play a key role in our portfolio strategy, the nature of these solutions will continue to tend towards fully-featured solutions, designed to run as free-standing individual systems, i.e. as islands isolated from one another. Within their targeted niche, these vendors have been successful in taking a functional area, integrating and unifying data, processes, and functionality within it, and growing their niche over time.

  In order for us to improve the overall research endeavor at Northwestern, we need to do very much the same thing: provide a framework that enables unification and integration across data, processes, systems and functional areas.

When taking on this unifying framework task, we also make six more ancillary assumptions:

1. While there are key vendors in the research administrative systems market sector, there is significant volatility in this sector. The optimal architectural solution needs to accept this as a "given" and emphasize the ability to swap systems/vendors in and out when necessary. This has implications for the extent to which customization of vendor products should be permitted, and the ability of the vendor to take advantage of Northwestern's infrastructural framework.
2. To be successful, the architecture must scale within the institution. The technical approaches employed in this architecture must be compatible and inter-operable with overall Northwestern IT infrastructure, and the Northwestern IT infrastructure needs to stress its ability to interact with third-party vendors and partner institutions via standard methodologies and technologies. This technical environment must be able to rapidly incorporate new software components with minimal modification, and allow for local software modules to be efficiently deployed when needed to bridge gaps between major systems.
3. While a number of the processes surrounding research need to be centralized – e.g. proposal submission, award acceptance, financial accounting, compliance with human and animal subject regulations – many of the processes associated with research administration will be school-, department-, or subject- centric – e.g. determining research opportunities, identifying available resources, aligning with inter-institutional partners, and reviewing proposals. Whatever ecosystem is built will need to allow for both centralized and distributed types of processes in order to be effective.
4. More than ever before, conducting research is a collaborative process on an individual, and even an inter-institutional, level. There is a trend to more co-authored papers, and research questions are increasingly including topics that cross the traditional boundaries of subject areas. In order to be successful, the ecosystem of research administrative systems must plan for, and work within, a cross-institutional setting, and the ecosystem must assume an ever-changing context of collaboration as normal, not as an exception.
5. Ease of access to data for reporting and analytics to ensure compliance, operating efficiency and timely decision making is critical to the success of our research enterprise. The reporting infrastructure should facilitate self-service access to integrated data while ensuring ease of access, clear data definition, and intuitive navigation.
6. In order for this vision to be realized, decisions about individual systems should no longer be made in isolation from one another. The research endeavor is large, complex, and cross-organizational in nature. Aligning it across business units and even institutions necessitates that a local perspective needs to be supplanted by an enterprise perspective, enabled and fostered as part of the IT Governance framework.

In short, the vision contained in this document is for a set of systems that consciously functions as a portfolio of systems; is explicitly constructed to be aware of, and interact with, one another and the surrounding infrastructure within which they exist; is designed to be flexible to accommodate change on multiple levels; provides "unifying" tools/windows for both researchers and administrators; and is purposefully guided by an enterprise perspective and governance framework.

The Value Proposition

This new path will require a long-term commitment, built upon not only the avoidance of an unsustainable situation, but also by the opportunity for gains in efficiency, compliance, cost avoidance,

and even in the quality of the research being undertaken. Underlying many of the examples that follow is the reality that, ultimately, the more time a Principal Investigator must spend on administration, the less time there is available for research, the longer it takes to transform an idea for research into an actual project, and the slower the advances in science and knowledge will be. We also believe that in some instances, improvements in administrative processes will lead even more directly to improved research, e.g. as clinical trial participant selection processes get refined administratively, the populations available to, and included in, trials should become even more targeted and appropriate.

The following examples illustrate aspects of the value proposition associated with the approach recommended in this report.

Benefits

Mitigate Risk, Avoid Costs, Improve Management, and Enable More (and better) Research

Integrating data across research administrative systems, and building unifying applications on top of it, will reduce the labor associated with day-to-day research administration, and will provide access to information about key business questions that are currently difficult to impossible to obtain or track over time.

As an example of some of the challenges (and opportunities for a new approach) in this area, consider billing issues within clinical trials. All human subjects engaged in clinical trials must be monitored to ensure appropriate billing to Medicare. To do so, one must know:

- the study on which they are engaged
- the status of the IRB approval related to that study
- the personal information related to the patient
- the billing information for services rendered

All of this data exists amidst the Northwestern (and clinical affiliate) ecosystem of applications, but given the current *lack* of data cohesion, system integration, and an overriding commitment to an enterprise perspective via institutional governance, it is taxing to even get the data, let alone integrate it and monitor it for this critical area of compliance. Were this data to become more easily accessible for aggregation, better monitoring systems could be created faster, and for far less money, to track care and billing, and to enforce compliance with sponsor regulations.

Similarly, the limitations imposed by the lack of data and application integration can be seen when you look at the everyday administrative barriers to conducting research at Northwestern. For instance, when it comes to the grant lifecycle -- pre-award phases (idea generation, proposal creation, submittal, and review) / post-award project management and monitoring / project results and close-out -- even seasoned researchers often struggle to obtain the information they need to make decisions, or to know what is required of them at transitional points in the lifecycle. Junior faculty may have no idea where to begin.

Once a grant is approved and a project is in flight, Principal Investigators are ultimately responsible for the conduct of their program. As such, they need to ensure that people get paid, supplies are maintained, budgets are adhered to, and reports are filed in a timely manner. Even though a considerable amount of time and effort is spent insulating the researcher from administrative burdens with both people and systems, these simple tasks can still be very challenging. On a very basic level, the

same information often has to be entered into multiple systems rather than flowing behind the scenes from one system to another ("enter once, use many times"), and getting an overview of one's project – available balances, recent transactions - requires producing independent reports from disaggregated independent systems, which then must be physically viewed or recombined to glean the needed information.

Whether you're talking about processes connected to project proposals, research billing, or budget review, moving these processes online can offer real savings potential for labor and opportunity costs. For these processes to flow online, the information needs to flow automatically in real time behind the scenes. By connecting the systems and eliminating the need to print out forms/reports and process them manually, effort is reduced, the likelihood of errors is reduced, time between steps is reduced, and it's easier to guide the participants through the next steps of the process. Additionally, if the data about the process itself is retained via logging, these processes set up a virtuous cycle by easily providing a rich source of process-improvement data that is difficult to obtain from paper-based processes. The hospital's work applying Six Sigma techniques to key business processes is a clear example of how this can work, but many savings can be realized even without this level of structure and sophistication around business process management.

Finally, just as it is difficult for researchers to get answers to basic questions, departments, schools, and the University as a whole have trouble getting answers to basic business questions in order to make informed decisions. Questions as crucial as "how much unrecovered effort are my faculty contributing to their research" are notoriously difficult to answer because they involve data from multiple, unconnected data sources. Integrating and managing data across research administrative systems, and then providing useful lenses to view/parse this information, could allow units to make more informed decisions about:

- Productivity/space usage
- The impacts of contributed cost sharing
- Recruiting
- Retention
- Performance evaluation
- Resource allocation

Support Work that Bridges Institutional Boundaries

Inter-institutional collaborations are no longer the exception. They are the new normal, and impediments to collaboration at the institutional, project, and individual level are a growing opportunity cost. In order for these barriers to be removed securely and effectively, solutions need to be designed and built with an extensible, enterprise perspective at their root.

At Northwestern, one of the more obvious examples of cross-institutional collaboration has been the relations between the Feinberg School of Medicine, Northwestern Memorial Faculty Foundation, and Northwestern Memorial Hospital. Clinical projects, awarded to Northwestern University, are staffed by NMFF members, and utilize healthcare and diagnostic services at Northwestern Memorial. Despite this long history of inter-institutional collaboration, the data sets and systems managed by these three organizations have remained almost completely isolated from one another, and their lack of integration is regularly referenced as an inordinate barrier to work.

One example of the limits of current solutions for connecting these institutions comes from the administration of clinical trials research at Northwestern and its affiliates. The eIRB and InfoEd systems are important components of the process, but in and of themselves, they do not constitute sufficient tools to administer a clinical trails program, and the eNOTIS application has been developed by the Feinberg School and its affiliate partners to help bridge these gaps. These systems are connected, where they are connected at all, through a tissue of nightly data exports/imports and point-to-point views, which are brittle, partial, and sub-optimal solutions. They reflect the reality of local business units trying to solve local business needs within an enterprise structure that retards cross-institutional collaboration.

Instead, cross-institutional collaboration should be seen as a growing reality and opportunity, and fostering it should be seen as strategic priority. Local solutions to local needs will continue to be needed. This work should be done, however, in the context of and reaping the benefits of an enterprise infrastructure that facilitates that work and the connections it needs to the rest of the enterprise.

Similarly, as partnerships and collaborations are encouraged across more institutions, the need to easily integrate "remote" participants into Northwestern processes becomes both more difficult and more important. These new members of an expanded Northwestern community, will be, by definition, less familiar with Northwestern resources or ways of getting things done, and will be less able to find assistance and support close to their normal daily responsibilities. As the number of people in this situation grows, the more important it is to have online services, that work in real-time, that are accessible using partner institution's federated identities, and that are integrated into unifying web applications that are easy to find and use. This is true whether the expansion of community is due to the changing nature of healthcare, the growth of new partnerships with institutions such as Argonne National Laboratory, or simply the growth of research collaborations with other R1 universities or private companies or individuals.

Costs

While we believe that this path comes with significant benefits and will ultimately be the only successful path for improving research administration, it clearly does not come without costs. There are two sets of costs that will be incurred:

The first set of costs is tied to the resources needed to build the new interactivity, most of which are not related to IT-specific work. Realizing this alternative vision will require a commitment to several large initiatives, e.g.:

1. Data will need to be rethought. Master Data Elements need to be agreed upon in definition and location, and processes to adhere to these agreements need to be designed and built. Data access needs to be rethought to facilitate the flow of information across systems. An integrated data warehouse repository needs to be built with common dimensions, aggregations, institutional metrics and access rules based on affiliation and position. Without this effort, reducing duplicative data entry, doing cross-functional analytics, and implementing online workflows remain difficult to impossible tasks.
2. In order for data to remain integrated, and for processes to become automated, integrated, and real-time, we must commit to an underlying services architecture as our standard way of

providing access to data and keeping it integrated. This means that the framework for this architecture must be built into the enterprise infrastructure, and services must be built within all of our systems to both expose data and processes, and consume external data and requests.

3. Presentation layers – portals, dashboards, and reporting and analytic tools – need to be built to facilitate the use of systems, the flow of business processes, and views on the data and the processes. These layers help change data into enabling information.
4. Online workflows need to be designed and built in order for tasks to be completed more efficiently and transparently, and for data about their functioning to be made available for analysis. Moving business processes online, improving them, and automating them is one of the biggest sources of potential time and cost savings.
5. A more robust unifying Identity Management infrastructure (including Single Sign on, identity federation with partner institutions, and a detailed Person Attribute Service) needs to be designed and built, and the portfolio of systems within the ecosystem (inside Northwestern and at our partners and affiliates) need to be modified to take advantage of this infrastructure when they make access and authorization decisions for their systems.

Undertaking these initiatives will require a large amount of work from a wide variety of units at Northwestern and ultimately our partner institutions, as well as a substantial financial investment. IT-specific work is needed in each of these initiatives, but the most difficult aspect of each one lies in the business work that must be made on processes, definitions, and policies. Addressing these topics effectively will disrupt business processes and distract from other priorities. If these efforts are to be successful, they need to be clearly prioritized within and across the institution.

The other set of costs will fall on the individual business units who "own" systems and are accustomed to having the flexibility to prioritize their own needs, investments, and timeframes based on their own narrowly defined priorities. These costs will come in two ways. First, as the enabling technologies and unifying applications are built out, individual line of business needs will have to be critically weighed against these cross-unit needs, and some will have to be delayed or reduced in scale if the enterprise vision is to succeed. Second, coordinated planning across business units will take longer to get agreement on priorities. This will be particularly true at the beginning, but as the new approach is learned and becomes the default, this time differential will go down.

The Opportunity of Now

The timing of this effort has never been better at Northwestern due to its alignment with a number of supporting efforts focused on building an enterprise portfolio of services:

1. The recent emphasis on Information Technology (IT) Governance and IT service federation prioritize coordination and cooperation in a new way. Discussions that have not taken place before now have new forums in which to happen and a new context to support them.
2. As part of its emphasis on "enabling technologies", discussions/planning/prototyping having already begun on:
    - integrating applications via services instead of the export/import of batch data files,
    - emphasizing the need to improve the Identity Management infrastructure and move access management out of the Identity Management system to its more appropriate home in the surrounding applications,

- standardizing data and building services in order to enable online workflows, automated business processes, and cross-function analytics,
- promoting a Single Sign-on infrastructure in order to enable application portals.

3. The most complex area of institutional differentiation is the University's relationship with the Northwestern Memorial Hospital and the faculty foundation. The last few years have seen important movement in the commitment to integrate these organizations and their attendant IT systems and infrastructure. The vision that is being proposed here is based on the same tenets on which those efforts are based. Coincidentally, the vision that is espoused here was also a defining element of NUCATS' recent CTSA funding proposal, so the basic premise of the need for unifying applications is already shared with a key part of the University research endeavor.
4. Interest has been expressed by WCAS, and been supported by other schools, for an idea that is sometimes referred to as a "PI Portal". The concept is really about having a one-stop shopping place for Principal Investigators and Business Administrators to get a unified view of data from independent systems about their research proposals and projects.

The discussions about collaboration and enabling technologies / unifying applications have started at the general enterprise level, and will help to define the context for the parallel conversations we hope will be happening in the research administration area. Our hope is that research administration can be a leading example in this effort across the enterprise. While research administration contains some of the biggest challenges in this discussion due to the inclusion of the FSM-NMFF-NMH dynamics, it also has the advantage of being a relatively cohesive line of business, and hopefully has opportunities for making advances that might be more difficult to realize in more diverse parts of the University.

Recommended Architecture

Ultimately, the goal at hand is to improve the business processes of research administration. In order to do this, the architecture that is envisioned must:

1. support consistent and documented compliance with grant regulations and reporting requirements,
2. be secure and collect access information as needed to monitor use of applications and data,
3. be aligned with the University's overall IT infrastructure so that it may rely upon a wide base of support and protection against obsolescence,
4. support distributed areas of management activity. Any assumption of a unified approach must be strongly justified due to the inconvenience coordinated adoption will impose upon all units/institutions,
5. be flexible enough to permit new business needs to be addressed efficiently and effectively through either local software development or introduction of commercial software without undue cost, and
6. be designed from the start to anticipate a mobile workforce that regularly will be outside the institutional firewall and operating from a public network.

The architecture that is being envisioned to meet these requirements has three main components, as illustrated in the diagram below (Figure 1).

At the center of the architecture is a consciously chosen and managed portfolio of transactional systems at Northwestern, along with transactional systems and/or data warehouses at partner institutions. Their value, minimized when left in isolation from one another, is optimized by the two sets of architectural components that surround and transform them: a consistent infrastructure and a set of unifying applications.

The consistent infrastructure insures security and provides the underpinning for the layer of "unifying applications" that bring cohesion to the collection of systems and functionality. Each of these architectural elements is briefly described below, and many of them are further amplified in corresponding Appendices.

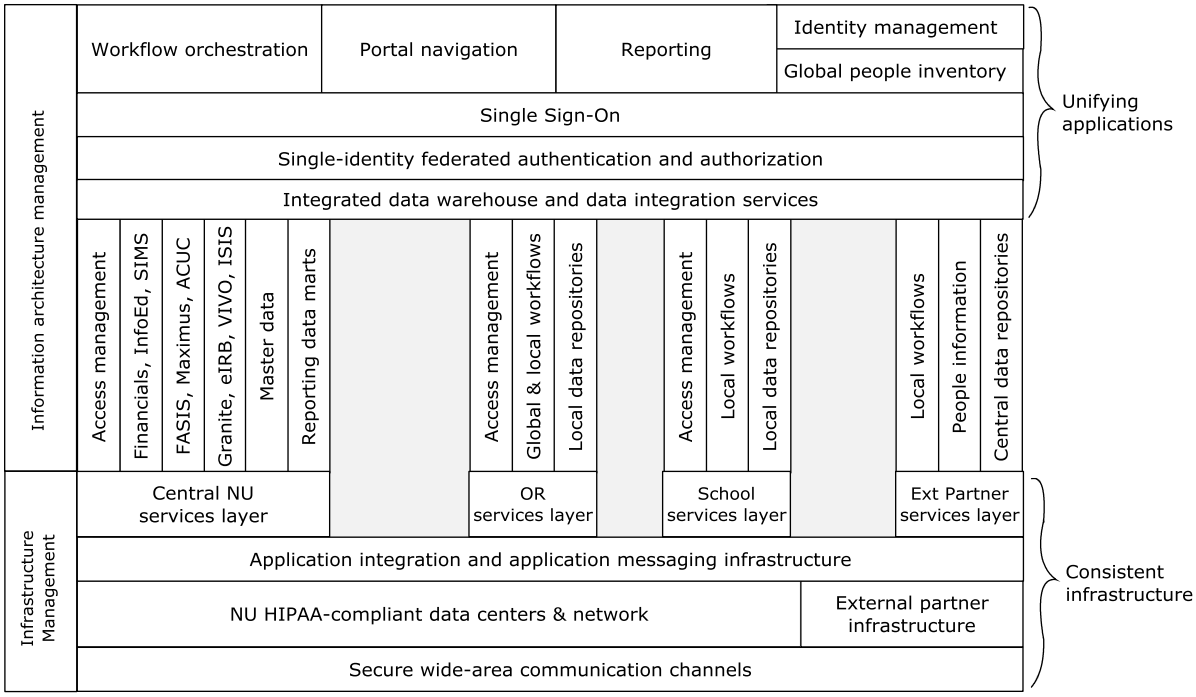


Figure 1 - Research Administration Logical Architecture

A Portfolio of Transactional Systems

The first component in the recommended architecture is the set of *transactional systems* used to conduct the many functions of research: creating the necessary records in systems and updating/maintaining them as changes occur throughout the research life cycle (prior to award, during the research performance period, and closing out the research). (Appendix A, page 26, has a list of the main systems used in research administration.)

Obviously, these systems play a key role in research administration, and if these vendor products are not

done well, the research administration experience suffers. When we look at these systems, there are many dimensions where we need optimized functionality. For example:

- their User Interfaces (e.g. their organization of information, the relation of screens to one another, the number of clicks it takes to get the desired information)
- the processes they present to the users (including data input, information lookup, and approval)
- their ability to be used easily and effectively on multiple devices (e.g. the user interface should be "reflexive" so that it is automatically optimized for whatever the screen size is of the device being used to view it; web browser interfaces are preferred over having to install software on a device in order to access the system)

These are just examples. Giving a detailed overview or listing of what should be included when evaluating one solution against another is beyond the scope of this paper. (A sample of a system selection tool that can be used in vendor selection processes is included as Appendix B: Solution Selection Guidelines and Tools, page 32.)

Clearly, however, in order for this vision to be realized, there are additional requirements for optimizing these third-party transactional systems within our ecosystem beyond them simply being good within themselves.

- First, as referenced in the initial assumptions of the vision, given the amount of work that needs to be done overall and the volatility in the marketplace, we need to be able to maintain and sustain these systems easily, or even swap them out if necessary.

  The biggest obstacle to this objective is having a vendor solution that is highly customized to extend functionality beyond the core product specifically for the Northwestern instance. In the short run, these customizations make for a seemingly great solution. In the long run, they consume valuable time, hours, and dollars when they break, when normal vendor software releases require their close testing and/or remediation, and when it comes time to swap a system out and all the corollary functionality that has been built into them needs to be rethought and re-architected. Each time one of these customizations is considered, there should an overridingly compelling impetus, there should be serious considerations of less invasive alternatives, and they should be minimized everywhere possible.
- Second, they also need to integrate with each other via Northwestern's infrastructural framework - e.g. be able to participate fully and easily within the services architecture, and interact well with the University's Identity and Access Management systems – and these functionalities need to be prioritized in vendor assessments. Appendix C: Architecture Requirements for System RFPs (page 33) translates the following sections into architectural requirements for System RFPs.

Consistent infrastructure

The foundation on which the transactional systems rest, and on which the unifying applications depend, is what we term a "consistent infrastructure". This encompasses physical facilities, security practices, and standards-based services for application integration. Northwestern, its partner institutions and collaborating institutions will have their own enterprise-scale infrastructures with these characteristics

as core to their other business functions, but these instances must work seamlessly together to support the research administration goals we describe. HIPPA/HITECH-compliant data centers, encrypted and firewalled communication links, and secure services architecture are pre-requisites to efficient cooperation between institutions and achieving those goals. Workflows span institutional boundaries and research data passes between institutions today - that will not change under the goals set herein - but those boundaries must not be high barriers to integrated applications that will efficiently support administration.

All facilities, networks, technology platforms, and operational practices must be compliant with any and all mandated requirements. Partner institutions will work together to establish compliant interconnections and operating agreements to preserve this environment. Individual research projects and stewards of data repositories will have formal working agreements for the handling of data elements, including encryption where warranted. Compliant practices are also required within individual laboratories and offices, which is the responsibility of the researchers themselves.

One of the biggest prerequisites for the unifying applications is the services infrastructure, whose global registry will be the management and functional location of services intentionally "exposed" in order to enable data reporting across units and institutions, global workflow support, and portal-based interactive gateways into authoritative systems. There are two components to this software integration layer that must be imposed to make it effective and consistent:

- Its core nature must be **real-time**. Self-service processes, and event-driven services (e.g. action 1 happens in system A, and then actions 2…x need to happen in systems B,…,Y, or person Z needs to be notified to take action D) are foundational to maintaining data integration, security, and putting processes online. This quite directly indicates that a services-based architecture, combining SOAP/SOA and REST/API approaches, will be appropriate.
- Services will be required from several levels of the University and its external partners, and everyone will need to retool their applications to use them as the default method of data exchange and access. For instance, central information and transaction systems must expose (i.e. make available for usage by appropriate systems/people) certain data and functions (e.g. financial and employment). The Office for Research must expose workflow functions for invocation by global and school workflows. Schools and external partners must expose similar workflow linkages.

Unifying Applications

The transactional systems in Northwestern's portfolio will usually be designed by their vendors as inwardly-focused, self-contained systems that provide a suite of functionality for a particular niche within the research administration process. The systems and data repositories at our partner institutions will similarly not be designed as explicit parts of our service infrastructure.

Our challenge, and the value that needs to be added in order to optimize the overall research process, is to unify these systems into smoothly functioning services that span functional niches, departments, and even institutions. This is the role of the **"unifying applications"** that sit on top of the transactional systems:

1. Identity and Access Management

2. Data Integration
3. Workflows
4. Unifying Views: Reports, Analytics, Portal

Utilizing the consistent infrastructural base, these "applications" abstract the separate institutional data repositories, tools, and workflows into a single presenting, reporting, and workflow environment, all of which is managed by a unifying Identity and Access Management framework.

The following sections give overviews of the role and defining elements of each of these.

1. Identity and Access Management

A key mechanism for integrating systems and services is a set of applications that verify and manage identities and authorizations for access to systems and data. While a full discussion of Identity and Access Management is beyond the scope of this paper, it is important to note that the primary organizing concept here is that each participant in the administration of research has a single identifier, and that the core component of this part of the infrastructure is a Person Attribute Service that will include all persons related to research, and the ability for systems to go to it and get the criteria they need to authorize what individuals can do or view within systems.

Person Attribute Service: This service would provide information about individuals that was available by agreement from authoritative sources. This service could be part of a University-wide service or be limited to just those individuals connected to research, both within the University and collaborators in partner and external institutions. Using information provided by the service, systems would make authorization decisions to allow or disallow access to functions or views of data. Appendix D: Person Attribute Service on page 37 contains a diagram of how this service would be structured and how it would interact with authoritative sources and various systems. It will need to be all inclusive, have the capacity to record multiple affiliations and job titles, and have a complete history of a person's' relationships to Northwestern.

Single Identifier: It is essential that each unique person participating in the workings of the research infrastructure be known within it by one and only one *identifier*. This is required for reliable access management, access auditing and reporting (particularly across transactional systems), and aggregation of data elements to individual researchers or research groups.

Built around this single identifier will be a set of re-enforcing ancillary functionalities:

- Any individual *identifier* may have **multiple *credentials*** associated with it (e.g. NetID, fingerprint, acceptance of 3rd-party credentials), but each person must have one identifier that uniquely identifies him/her.
- In addition to the important compliance benefits that come from having a single identifier per person, it also enables a much appreciated customer service function: **single sign on** (SSO). With SSO, one authentication action in any portion of the research administration architecture should be sufficient to access all transactional systems and unifying applications, and it allows systems to easily aggregate content from different systems for any person. All cross-system workflows and portals rely upon single sign-on technology.

These pieces will be key elements of an integrated Identity and Access Management system which will enable a single-identifier, federated authentication service. This service will operate at the center of the distributed environment that includes Northwestern, our partner institutions and collaborators, and it will allow convenient - yet secure - access by all participants using credentials issued by their home institution, and it can allow authoritative information to reside in-place, while simultaneously allowing overrides to be put in place locally where important.

2. Data Integration

The second unifying "application" is an information management architecture focused on making sure that the data itself is integrated and optimized. This unifying application is included in the logical architectural diagram on page 10 as the "Information Architecture Management" vertical bar on the left, spanning both the transactional systems and the Unifying Applications. It is more of a procedural "application" than a technological one, but its existence is critical to every other unifying application: the more fragmented the data is, the more duplicative and out-of-sync it will tend to become, thereby not only wasting valuable effort to maintain it, but also undermining any effort at cross-functional analytics or workflows.

Much of the data integration effort area needs to be on mapping the existence of data in the various systems, agreeing on shared definitions and authoritative sources of data, and setting standards for data usage. However, an effective data integration solution involves more than defining and standardizing data elements and usage. There are two more important ways in which a data integration approach needs to add value.

First, in order for data to remain integrated, it needs to easily flow between systems or be accessible when needed by another system. An example of the first instance is that currently the Principal Investigator for a sponsored research project is initially entered into InfoEd as the proposal is being created, and then sent via a regular data feed to NU Financials if the proposal is funded. However, should there be a change during the course of the project in who the PI is, none of the subsequent changes made in InfoEd ripple through to NUFinancials because they are changes to existing records, not the creation of a new record. The authorization system described in the preceding section is an example of the second instance. Each transactional system in the portfolio should not have to keep track of each person's role on the project, and people should not have to re-enter data that has already been entered in a system. All these systems should need to know is: if a person has project role X, they will be able to do Y. Then, when a person requests access to Y, the system can query the Person Attribute Service, use it to find the information it needs to know about the requesting person, and then, based on that information, grant or not grant access to Y.

Each of these aspects of data "flow" depend on a robust services architecture, which will need to operate within a consciously designed set of rules designed to ensure that master data is never overwritten by ancillary duplicate data (unless explicitly done so by design), and is always available when needed.

Data warehouses illustrate a second way that integrated data can add value: via the aggregation of data. In data warehouses, data is specifically re-organized and re-structured into a unified repository to optimize analysis and reporting instead of transactional processing. A similar idea is also at work in the Person Attribute Service, that aggregates data (albeit virtually), so that requesting systems do not have

to keep track of where their needed data is, nor does it have to be duplicated anywhere. All a system needs to know is that when they need to make a decision on an access request, they invoke the service with a request for attributes, and the service gets the attributes that are needed from where they naturally reside.

See also Appendix E: Data Integration (on page 39) for discussions on data dictionaries, master data management, and data warehousing, and the discussion of data governance in the Governance section below.

3. Workflows

The third set of unifying applications is workflows. The unifying applications mentioned thus far are all exceedingly valuable, but in many ways they are only the precursors to the real value that is available via the ability to implement workflows.

Very few, if any, online workflows currently exist other than the workflows that are natively built into the transactional systems we have. These workflows are built into their internal code base, and to the extent those workflows are able to be "exposed" to external systems, they can be further leveraged outside of those systems. However, the real need we have is to build workflows outside of, in and out and between, these transactional systems.

The need for workflows comes in a variety of contexts. One has already been mentioned: the value of a data element gets changed in one system, and it needs to ripple through other systems. Either the new value needs to be passed to other systems for updating of their instance of the element, or the new value needs to be passed to other systems so appropriate changes to dependent elements in their system can be changed. These are ***automated event-driven workflows that happen behind the scenes***.

There are also ***workflows that guide people through processes***. For example, Principal Investigators (PIs) need to approve their monthly expenses. Currently, a report needs to be generated, printed out, walked around to the PIs by a research administrator, physically viewed and signed by each PI, and then logged and filed by the administrator. All of this could be automated online with no printing or walking: the report becomes available with the PI and the administrator notified when it is ready; the PI reviews the report and either accepts it or enters questions for review, either of which would automatically generate alerts for the administrator responsible for this process.

On a grander scale, the guiding workflows can act as **expert systems** to assist researchers compile and organize information for a research proposal based upon granting agency, discipline, research approach, etc. The workflow "walks" the researcher from one department to the next to identify necessary information, request reviews and approvals, and ultimately submit a completed proposal without special knowledge of the route that was taken. More time is spent on a quality proposal and less on navigating the bureaucracy.

Workflows not only add value by unifying applications via smooth real-time data flows, and by facilitating complex business processes, they also add value because information about the execution of the workflows can be captured via behind-the-scenes transaction logging, which can be analyzed subsequently to improve the business process.

What is being envisioned here is a top-level workflow system to drive overarching, cross-institutional

business processes. Top-level workflows will trigger other workflows as they progress. While a monolithic structure could have simplifying features, it should be possible to break down long flows into shorter segments that can be started in turn based upon hand-offs between responsible units. This will allow local workflows to be authored and maintained in simpler (or, if needed, more feature rich!) software than the central engine.

The value of workflows was acknowledged in the recent Clinical and Translational Sciences Association funding proposal via its emphasis on NITRO – a workflow engine to tie the research-related systems together:

"The major innovation of NITRO will be the incorporation of a robust workflow engine that supports the Business Process Execution Language (BPEL)."

This working group shares the NUCATS assessment of the value to be gained from implementing online workflows, and it is recommended that opportunities for an infrastructural synergy be sought out and purposed in this area.

For more detail on how workflows can be leveraged in existing systems, and an example of how workflows could be employed to improve the process of research proposal creation, see Appendix F: Workflows – Using Data to Manage the Processes of Research on page 41.

4. Unifying Entry/Views: Portal, Reports/Analytics

While there is much to be gained from improving how the data and business processes flow within the research administration endeavor, there is also great value available in improving the doorways and windows into both the data and the systems that manage it. Improving these "views" can make it qualitatively easier to use the wide-ranging set of research administration systems, can bring much higher levels of efficiency to the business processes of research administration, and can also help drive improved decision-making.

Portal

One of the most common complaints about the process of research administration at Northwestern is that it is so vast, fragmented, and difficult to know. Whether it is a new researcher or a veteran research administrator, everyone talks about this. Providing a unified front end to the array of research administration systems and processes at Northwestern should be an important goal.

Portals organize entry points into what would otherwise be a confusing set of disparate systems, and when done well they can also be a personalized "home base" for getting information about / views into an otherwise disparate set of processes. What is envisioned in this report is a feature-rich portal navigation system that will be the recognized, unified view into all work in progress related to grants, from conception to close-out.

Examples of what would be included in this portal include:

- alert information about processes occurring in or across multiple systems, which can be "exposed" within the portal and consolidated in a personalized Alert Inbox
- dashboards and simple reports to help each user monitor projects and processes in which they are personally invested

- workflow wizards to help shepherd people through common business processes, particularly those which span multiple systems

In all of these cases, the solution being proposed depends on the architecture described heretofore, e.g. a robust service architecture, single sign on, data integration, robust but flexible security mechanisms, and ease of use considerations such as browser interoperability and support for access from mobile devices.

Like the other aspects of this architecture, any portal view into research administration should be part of the University's portal. There are two overriding reasons for this: providing one unified portal consolidates the research side of people's jobs with the other online aspects of their lives at the University that are manifested in the other parts of the portal, and the consolidation optimizes the portal's chance of success by creating a critical mass of entry paths to systems/processes/information.

Reports/Analytics

Reports bring data sets and elements together so they can be viewed in relation to one another. They can be all-important for both day-to-day operations and larger management overviews, yet they're often overlooked when systems are being considered or resources are being allocated. 3rd-party systems usually come with a standard set of reports, but they're limited to the host application, and typically they are very "bare bones" and need to be supplemented by reports designed and built locally to best reflect local business processes. No system is going to come with reports that provide views into processes and data that cross systems.

There are numerous opportunities throughout the lifecycle of a research grant/contract to improve the research process by providing "reports" to researchers, their teams, and their administrative staff. Often, the opportunity is not thought of as "a report". For instance, dashboards and simple reports should be presented at the portal layer; additional elements should be embedded into the transactional systems where relevant; more in-depth reporting should be available in a single click from either of these sources.

Much of this reporting functionality is dependent on having integrated data. See the section above on Integrated Data, and Appendix F: Data Integration on page 39.

For more detail on portals and reports, see Appendix G: Unifying Views – Portal, Reports/Analytics on page 46.

Governance - Organizing Unification

In order to change the inventory of research administrative systems into a portfolio of systems strategically built on an enterprise level to span institutions and provide end-to-end integrated online services that span multiple systems, conscious coordination must replace the more narrowly-focused line-of-business orientations of today. This coordination, i.e. governance, is the last "unifying application", which is a people and process "application". It will need to take place on two levels: the service portfolio level and (as already referenced in the Data Integration section and its related appendix) the data level.

Service Portfolio Governance

There needs to be one overarching governance committee for research administration, integrated within the current IT Governance (ITGov) framework, and focused on promoting and ensuring this enterprise portfolio approach. The representatives on the committee and the agenda it follows should align with the inventory of research administrative systems in Appendix A on page 26.

The predominant reporting relationship within the ITGov framework probably should be to the Administrative Systems Advisory Committee (ASAC), but it should also have a connection to the Research Technology Advisory Committee (RTAC) once that committee is launched. The committee will provide oversight over a portfolio of research administrative systems and projects, ensure alignment between investment in research systems and institutional priorities, and make recommendations on roadmaps and future investment in systems. The committee should have representation from functional areas within Office for Research as well as key research intensive schools including the Feinberg School of Medicine, Weinberg College of Arts and Sciences, and the McCormick School of Engineering.

Data Governance and Stewardship

Historically at the University, information within each system has been governed by a data steward, who is chartered with ensuring (1) that their system is able to support a particular set of business functions (often fairly narrowly defined to its line of business), and (2) that the system's security policies enable appropriate access to the data within the system.

To encourage and insure the success of an enterprise-wide portfolio of services, barriers to data integration must be removed. Those barriers may arise from administrative priorities, security concerns, required technical efforts, or mis-alignments of data element semantics. Each of these possible sources should be addressed and minimized wherever possible to establish an environment where data about a single entity (person, project, grant, financial account) can be assembled for workflow, analysis or reporting with minimal effort.

Within the research administration area, data types such as "protocol" and "proposal" have specific definitions and can be governed within the Office for Research. Other relevant data reside in various systems outside of the Office for Research governance sphere, such as personal data for a Principal Investigator, engaged graduate students, or assigned physical space. The recommended architecture in this paper depends heavily upon lowering the barriers to incorporating relevant administrative grant data from multiple systems into interconnected data sets that would preserve a record of the grant's history and current status.

A standing committee on Research Administrative Data Governance should be assigned the role of ensuring that data definitions, semantics, and privacy classifications are consistent across the research enterprise. Partner institutions should be represented. This might be a subcommittee of a larger University-wide committee with a similar charge for all administrative and academic data. That committee should also be charged to define consistent, reasonable, and readily implemented security methods for making data items available electronically as needed.

Recommended Next Steps

Overarching Changes

This working group recognizes that its recommendations are ambitious, yet it believes that incremental changes within the current model of providing research administrative services will ultimately be unsustainable and self-defeating within expected resources as the scope and pressures on research administration grow at their forecasted rates. The administrative environment must be overhauled, integrated, and streamlined to create efficiencies and value while preserving local management control over key functions as they are defined today.

We also realize that the recommendations that follow, and are contained in, for example, Appendix C, are mostly categorical in nature, while we live in a non-categorical world and the real life choices we will face will often not be so black and white. We recognize this disconnect, but we intentionally "overlook" it to make a strong statement encouraging this new approach to development.

The sections below list steps that need to be taken in each of the areas reviewed in this report in order to bring about the adoption of the recommended architecture. Some of the steps have core enterprise infrastructure prerequisites. Many, if not most, require concentrated involvement of business (as opposed to "IT") staff.

Transactional Systems

1. The practice of customizing 3rd-party COTS solutions to extend the vendor's functionality or match Northwestern business processes should be curtailed to the greatest extent possible. Inserting custom code, or repurposing native fields in the database schema for non-native purposes, makes on-going maintenance, upgrades, and system replacement qualitatively more difficult and expensive, and should be resisted.
2. As delineated, in most of the sections below, when choosing between alternative solutions in a functional area, each solution's ability to integrate within the Northwestern enterprise infrastructure should be prioritized.

Service Architecture

1. Design and implement the core enterprise services architecture. It is a precursor to the majority of this vision.
2. Develop service architecture expertise within ORIS.
3. Assess the capacity of existing vendor and system capability to expose and ingest data via a service architecture.
4. When assessing any potential vendor or system, prioritize the ability of the solution, irrespective of its programming techniques or platforms, to integrate within a robust services architecture.

Identity/Contact Management

1. Define a "global people inventory" (the basic part of the Person Attribute Service) of all persons who are known to the research administrative function – including staff, faculty, researchers, post-docs, graduate students, affiliates, undergraduate students, medical residents, technicians, etc.

2. Within this global people inventory, define a single identifier to be used by a person in all systems, tied via, federated services to verify the ongoing standing of persons as asserted by their home institution
    a. The resulting unified identity infrastructure will span the participating organizations such that, for the Research Admin IT ecosystem, a single person will only use one identifier for all work - that identifier will be valid for access by that person from any location
    b. Any single identifier will only be used by the designated owner of that identifier
3. Develop the ability within the enterprise infrastructure for providing multiple ways to integrate systems into the University's Identity and Access Management infrastructure.
4. When assessing any potential vendor or system, prioritize the ability of the solution, irrespective of its programming techniques or platforms, to integrate within a standards-based enterprise Identity and Access Management architecture.

Information Architecture

1. Document the information architecture across all systems and identify where data has been replicated.
2. Develop roadmaps for modifying:
    a. Data structures to use common identifiers throughout the underlying data repositories
    b. Systems to access data when needed via services to the Master Data Element holder
    c. Existing data mart silos so they share common dimensions and are integrated into a unified data warehouse repository with appropriate data classification and access provisioning based on affiliations and positions
3. Create a research project identifier that can be used to organize and refer information back to the original research project and proposal for analytical purposes. Assign this identifier at the inception of the research project, regardless of funding outcome.
4. Continue to link individual proposal and protocol records in InfoEd and Click systems to permit extraction to an integrated data warehouse.
5. When assessing any potential vendor or system, prioritize the ability of the solution, irrespective of its programming techniques or platforms, to integrate within the enterprise information architecture. For instance, in the case of proposal and protocol systems, they must accept an outside parent research project identifier initiating, and linking, related proposal and protocol submissions.

Workflow implementation

1. Adopt a central workflow technology for research administration, and build flows from the top down to deliver early value to researchers to build proposals.
2. Clearly state to all schools and affiliates that, by a given future date, all research administrative tasks will require their support through local workflows that will be invoked by the top-level flow.
3. Mobilize interested IT units on campus to help look at potential workflow solutions. The likelihood is that we will not find a one-size-fits-all engine, but will instead end up with at least two: a higher-level engine that is very fully-featured but requires special skills to work within it,

and a more distributable, rapid application development tool that can be used locally to build forms and workflows.

4. Look for opportunities (where there is a convergence of a pain point, a champion, and a favorable software application context) to implement local workflows. Where possible, link them via service calls to the top-level workflows.
5. Introduce a workflow on top of the Click IRB system allowing the study coordinator to enter study-related but not protocol-required data needed by FSOM clinical trial systems. This will permit a "vanilla" eIRB implementation while maintaining FSOM operations. This workflow could initially be provided by Click and later by an enterprise workflow architecture.
6. When assessing any potential vendor or system, prioritize the ability of the solution, irrespective of its programming techniques or platforms, to integrate within an external workflow orchestration system.

Unifying Views: Portal and Reporting/Analytics

1. Identify requirements, architecture, scope, technologies, and roadmap for a research Portal that will provide a unifying access point for both workflows and reporting; identify and engage with an executive sponsor to champion the project.
2. When assessing any potential vendor or system, prioritize the ability of the solution, irrespective of its programming techniques or platforms, to integrate within our enterprise reporting/analytics framework.

Governance

Create the two IT Governance groups: Research Administration Service Portfolio and Research Administration Data Governance and Stewardship.

Recommendations for Upgrading the eIRB System

The Office for Research plans to upgrade the current Huron Click eIRB system, and the working group was requested to have recommendations for how this upgrade should align with the vision contained herein. Over the years, the existing eIRB product has been extended out to connect with related functional areas internal and external to Northwestern (e.g. FSM, NMH, eNotis). Further additional connections are highly desired (e.g. NU/FSM COI, CITI, and external IRBs from institutions such as VA and LCH).

The eIRB project goals that were shared with the working group at its outset were to have a resulting solution based on a version of the current product's latest release that is "as vanilla as possible", while improving upon the current methods of information exchange between these key areas and partners. The end result should be feature rich but more efficient, and more maintainable and sustainable.

The section below - particularly when read in context of the architectural recommendations in the preceding section and detailed in the Appendices, particularly Appendices B and C – contains the key recommendations for this specific upgrade project:

Fundamental Tenants for eIRB Upgrade

1. Minimize to the fullest extent possible any "Northwestern-only" code changes to the core vendor product. If changes are made to the vendor product, the overriding preference is to

have them made by the vendor in their standard production codebase instead of as a custom Northwestern solution outside the regular product upgrade path.

2. Commit that no schema changes will be made to the core product's databases.
3. Commit that the default method of integration of the vendor product into the Northwestern eco-system will be in real-time via re-usable Web Services (either SOA/SOAP or REST/API or both). Where this is not possible immediately, paths towards this behavior should be mapped out.

eIRB Functionality Review

Before proceeding to the upgrade, OR needs to inventory all current eIRB local business function customizations**.** For those functions that are not included as standard functionality in the proposed new solution, their criticality should be assessed by a group that includes not only the affected business unit, but also those responsible for the overall project (AVP for Research Compliance, project sponsor, etc.)

1. If the function is no longer needed or deemed too expensive for the overall project, it should be discarded.
2. For each remaining functional customization
    a. Identify how both the current eIRB data schema(s), and those included in the new release, organize data needed by the business function customization, and how data needs to flow between the relevant systems.
    b. Identify how Web Services can be used to read and write relevant data on a real-time basis within the product's databases.
        i. Use vendor native Web Services if available
        ii. Otherwise, write and test Web Service shims to access data elements
    c. Write a new satellite application to deliver the desired function outside of the vendor product using Web Services to read and write data.
    d. If necessary, define a satellite database to logically extend the data held in a vendor product database record.

Implementation Notes

1. This recommendation assumes that a secure Services Architecture is available to protect applications from unauthorized invocation of services.
2. It will be highly desirable to align the development tools and technologies used to build any supplemental applications with those used by other administrative applications outside of the PeopleSoft applications. This alignment would enlarge the pool of talented persons within the IT@NU community who can consult on best-practices as well as assist in troubleshooting.
3. Some business function customizations require user-interface design and testing. ORIS will need to pursue collaborative design approaches with the research and research administration communities.

Appendices

Appendix A: Overview of Research Administrative Processes, Systems, and Data Flows

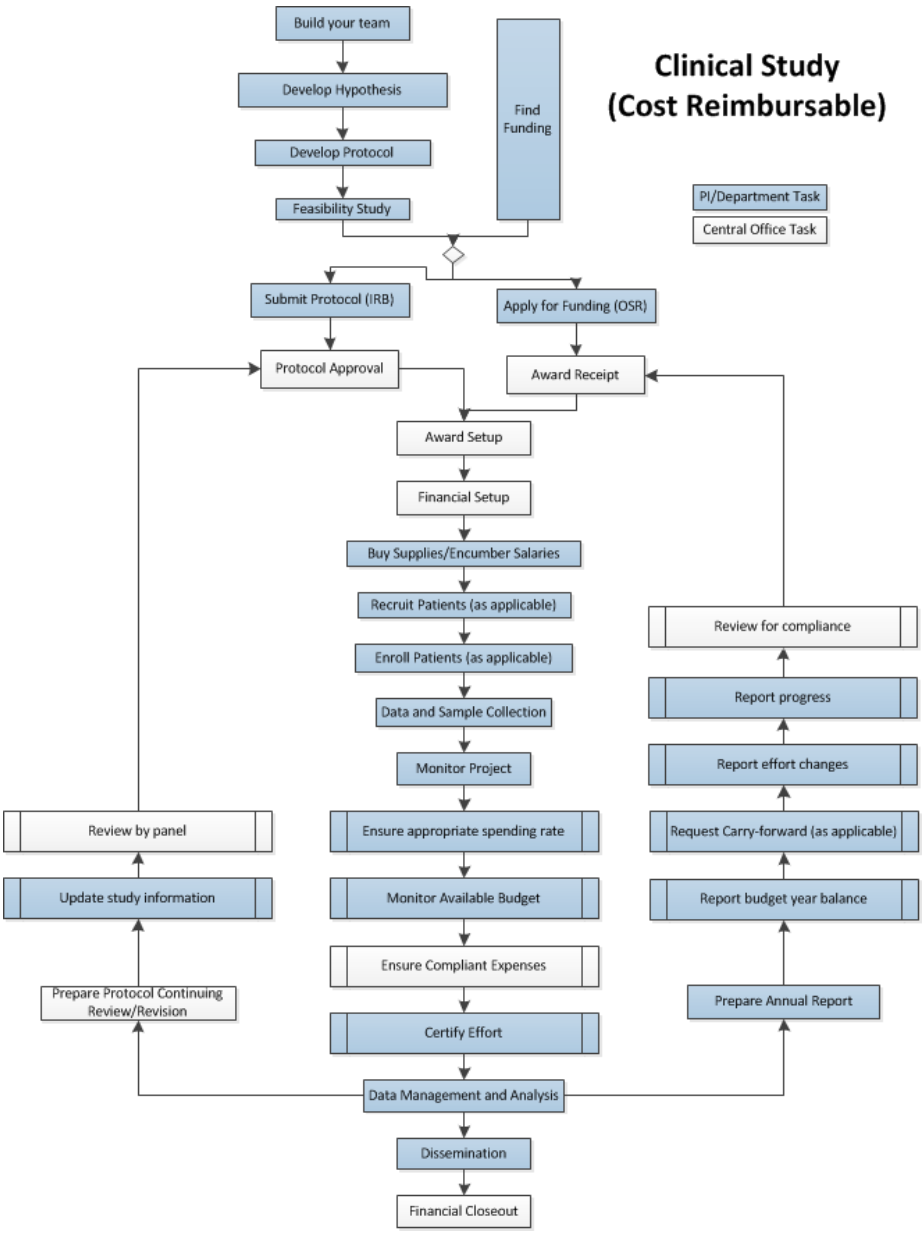
I. Primary administrative systems supporting the research process at Northwestern

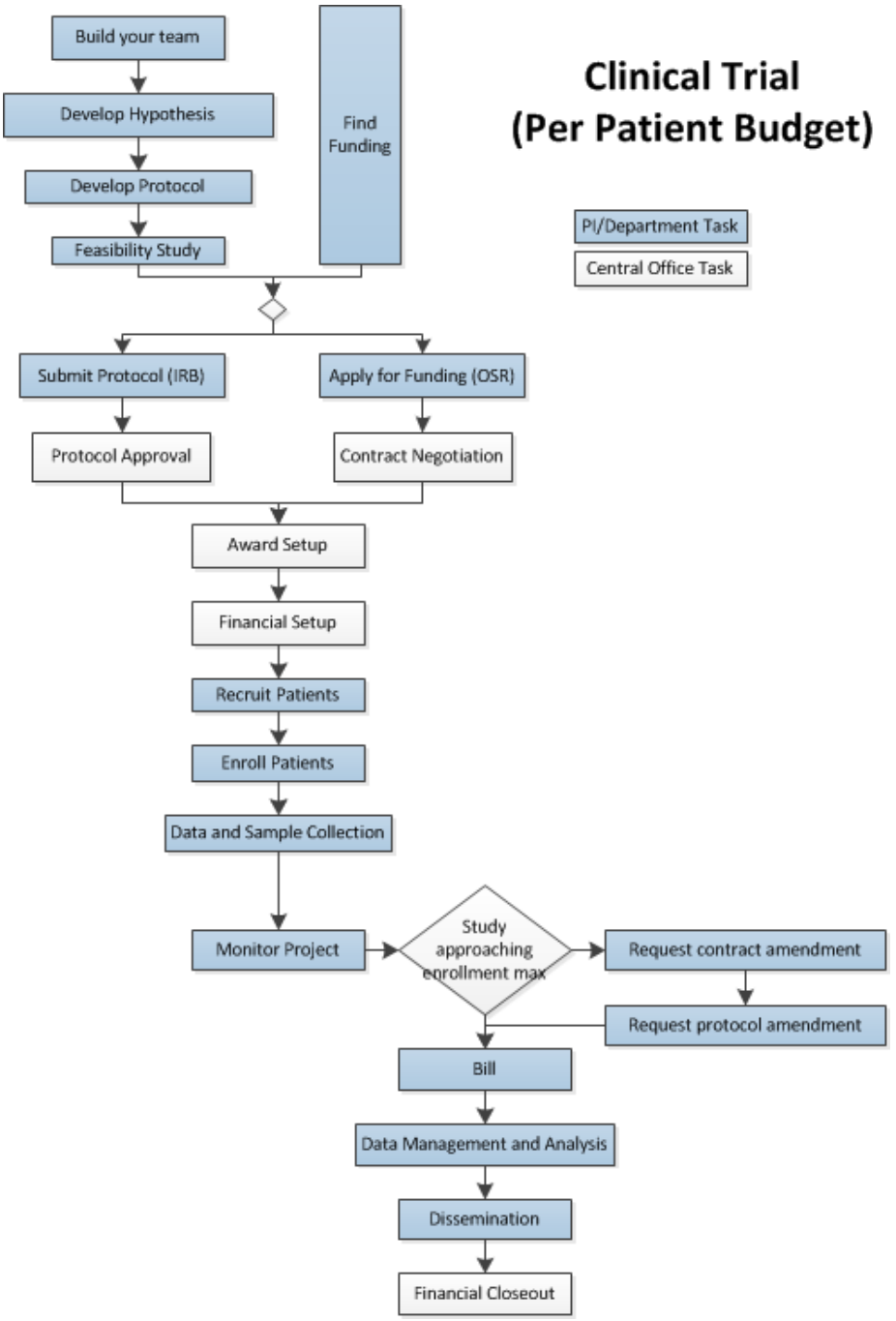
System	Vendor	Business Office	Description
InfoEd	InfoED Global	Office for Sponsored Research	Research proposal development and tracking
eACUC	Huron Click Commerce	Institutional Animal Care and Use	Animal subject protocol development and tracking
eIRB	Huron Click Commerce	Institutional Review Board	Human subject protocol development and tracking
NUFinancials	Oracle/Peoplesoft	All	Establish budget, track expenditures and encumbrances
Granite	Topaz	Center for Comparative Medicine	Animal facilities operations management
ISIS	ORIS Custom Application	Office for Research Safety	Lab safety management and training
Sophia	Wellspring Worldwide	Innovation and New Ventures Office	Technology commercialization management
ERS	Maximus	Cost Studies	Effort commitment and reporting management
FCOI	Oracle/PeopleSoft custom application	NU Conflict of Interest	Significant financial conflicts of interest management

The table above represents only the primary central office systems used to manage research administration at Northwestern. The central offices listed above utilize many additional satellite systems for specific tasks that are not accommodated by these primary systems, but these satellites serve a supplemental role, and do not contain authoritative data about the process.

This group recognizes that school-specific systems are often used to address local business needs and workflows. These systems vary widely in scope and support, and can be critical to the business functions in a segment of the University's business. NUCATS, for example, has built a suite of applications to address a number of needs in the Feinberg School of Medicine including RegiStar for clinical trial participant recruitment and eNOTIS for participant enrollment. These systems are not itemized in this document, but it is acknowledged that they will need to be considered in the implementation of the recommended approach.

II. Workflow Overviews of Research Administrative Processes


Clinical Study
(Cost Reimbursable)
PI/Department Task
Central Office Task
Build your team
Develop Hypothesis
Develop Protocol
Feasibility Study
Find Funding
Submit Protocol (IRB)
Apply for Funding (OSR)
Protocol Approval
Award Receipt
Award Setup
Financial Setup
Buy Supplies/Encumber Salaries
Recruit Patients (as applicable)
Enroll Patients (as applicable)
Data and Sample Collection
Monitor Project
Ensure appropriate spending rate
Monitor Available Budget
Ensure Compliant Expenses
Certify Effort
Data Management and Analysis
Dissemination
Financial Closeout
Review by panel
Update study information
Prepare Protocol Continuing Review/Revision
Review for compliance
Report progress
Report effort changes
Request Carry-forward (as applicable)
Report budget year balance
Prepare Annual Report

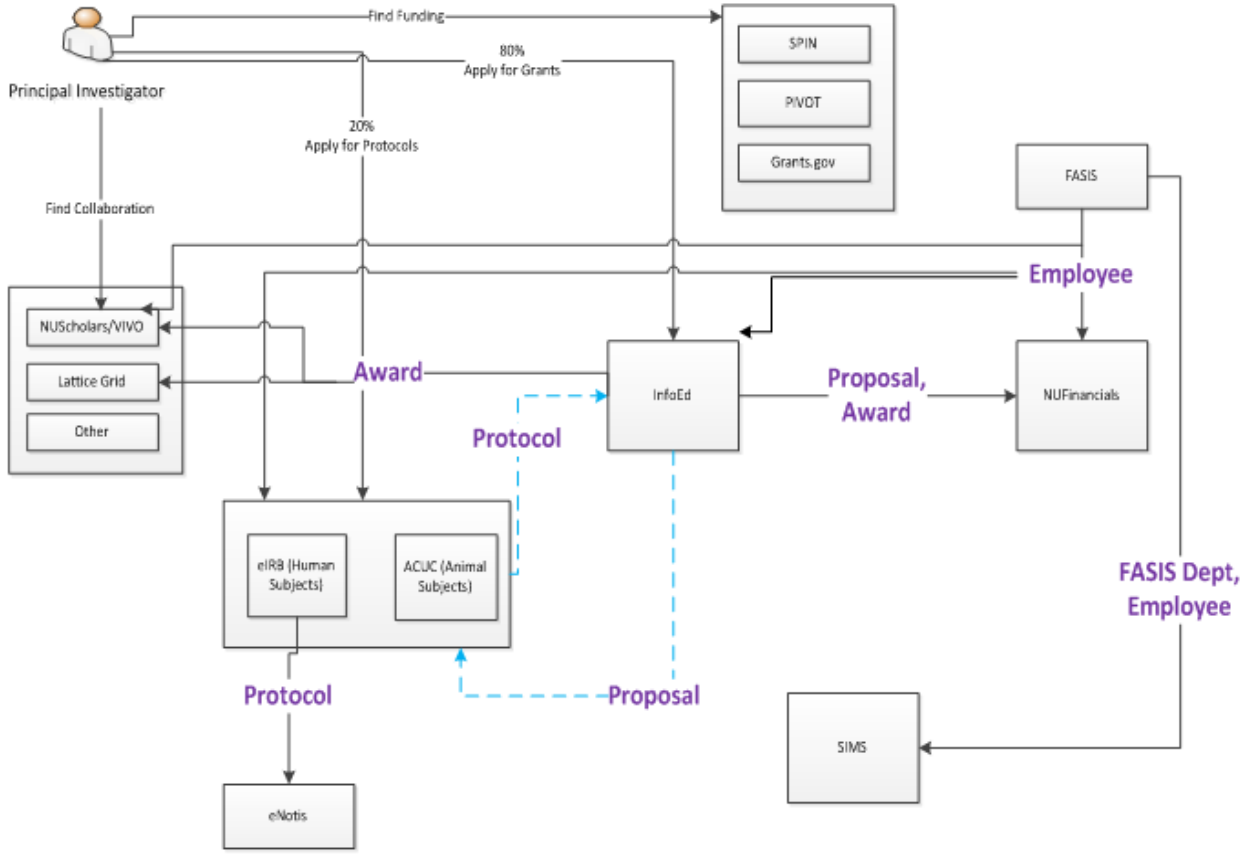


III. Diagrams of Data Flows between administrative systems across the research life cycle

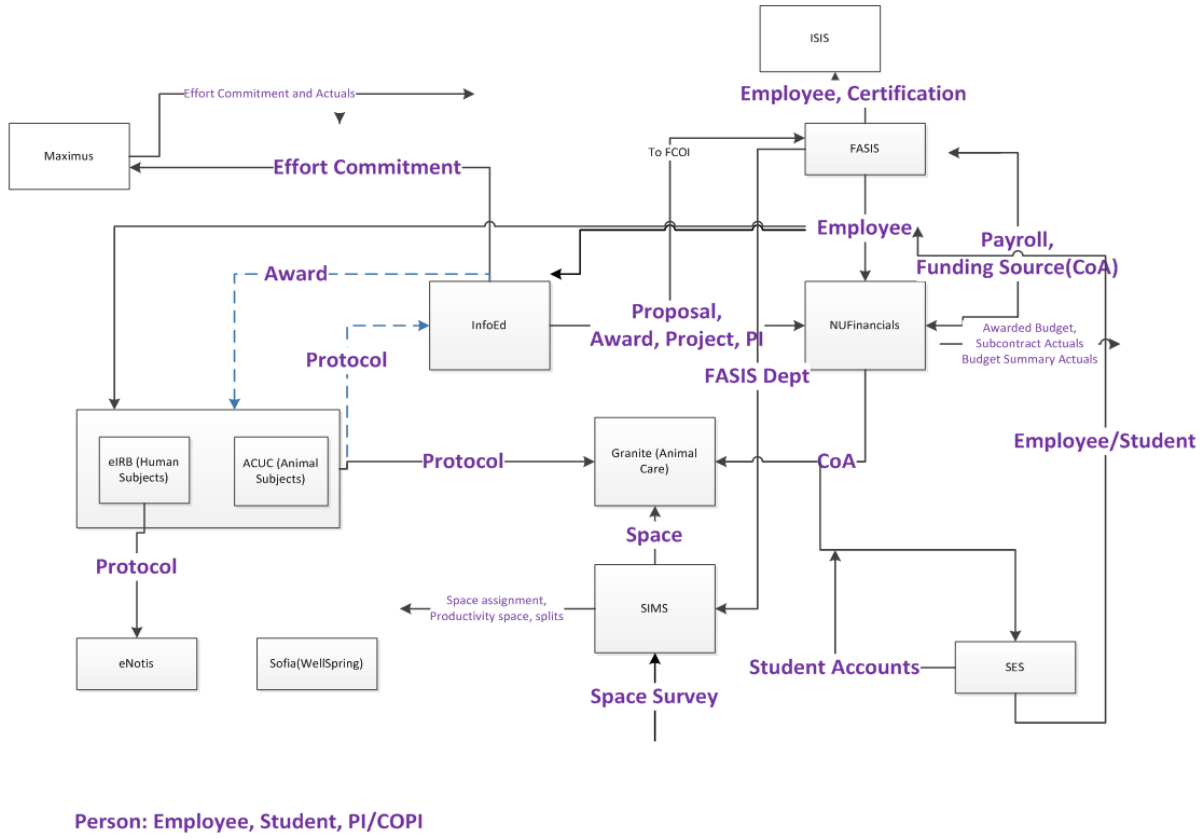
The ecosystem of research application systems, data interfaces, nightly loads and views has grown complex by accretion. The following figures give a sense of the complexity of the systems and the data exchanges that currently provide what integration exists between the systems.

The first diagram shows information flowing between research administrative systems (RAS) during the pre-award/proposal stage of the funding lifecycle. The second diagram shows information flowing between RAS systems during the post-award stage of the funding lifecycle.

i. Data Flow: Pre-Award to Award



ii. Data Flow: Post-award



iii. Cross-system Flows of Data Families

The following matrix is another view on how data flows between administrative systems.

- The blue rows contain sets of data that are utilized in more than one system.
- The red lines highlight instances where more than one system can, over the life of a grant, be considered to be the "authoritative" source of that data, leading to data fragmentation.

If the data in either of these types of situations does not flow easily and in a tightly coordinated fashion, the research process is impeded.

Data Family	InfoEd	NUFIN	Maximus ERS	Wellspring	FCOI	FASIS	eIRB	eIACUC	ISIS	Changes in state
Proposal	XX									
Demographic	XX									
Budget	XX									
Award	XX	X	X		X					
Demographic	XX	X	X		X					
Personnel	XX	X								Add/drop over time
Budget	X	XX								Add incremental, rebudget
Project	XX	X								Add incremental
Committed Effort	X		XX							update at increment
Certified Effort			XX							quarterly certification
Subcontract	X	X								
Demographic	XX									
Budget	XX	X								Add incremental
Agreement	XX									
Sponsor	X	X	X				X	X		Name changes, address, mergers
Subcontractor	X	X								Name changes, address, mergers
Department	InfoED	X	X	NUFIN	InfoED	FASIS	X	X	X	Move in tree, name changes
People	X	X	X	X	X	XX	X	X	X	termination, address, citizenship, name change
Appointment	X					XX				termination, addition,
Expense		XX								approval status, journals
Encumbrance		XX								
Invoice		XX								
Invention Filing				XX						patent status
COI Disclosure					XX					annual renewal
Protocol							XX	XX		periodic renewal
Demographic							XX	XX		
Personnel							XX	XX		add/drop
Scientific data							XX	XX		
Enrollment data							XX			amendments
COA Link							XX	XX		
InfoEd Link							XX			
Facility Data								XX		
Safety Plan									XX	
Demographic									X	
Personnel									X	
Room									X	

Legend

XX	Authoritative source of this data
X	Data utilized "downstream"
	Data Flow where data created in one system is needed in other systems
	Data Flow with multiple sources of "authoritative data"

Appendix B: Solution Selection Guidelines and Tools

With the increased volume of business systems offerings tailored to meet the needs of higher education business process, the recent proliferation of these 3rd party applications should be expected to continue. Many of these applications are offered in a hosted model and for a relatively low cost (both up-front and ongoing). While there are some risks inherent to increasing the number of products we are using and the adoption of hosted solutions, we are not suggesting that the use of 3rd party applications is a poor decision. Instead, we wish to point out that a strong process is needed to ensure that the solutions are selected, implemented, and supported in the robust manner necessary for the University to obtain the intended value. Several important aspects that this process should include are:

- **Efficiency** – the solutions chosen should be cost-effective and should not overlap with existing functionality in the University's enterprise systems (assuming such functionality either meets the University's needs or can be customized to do so with a reasonable effort level)

- **Integration** – the concept of a "standalone" system is essentially extinct; our assumption should be that we will be integrating these 3rd party systems with our core enterprise systems. Therefore, the systems chosen should be selected based on their ability to be integrated. Effort should also be allocated in our planning and prioritization efforts to tackle these integration projects

- **Security/Compliance** – in some cases, specific regulatory requirements must be met in these 3rd party applications in the same way that such mandates apply to our core enterprise systems. In all cases, the data stored in these applications must be protected against unauthorized access by both Northwestern users and external audiences

- **Support** – the solutions chosen should be accompanied with support from the vendor, including implementation assistance, training, customer support, and ongoing product development

As stated above, the desire to acquire, implement, and integrate these 3rd party applications is likely to increase. Bringing together groups such as ITMS, ISSC, and OGC to ensure that 3rd party system acquisitions are optimized across the enterprise is good practice. Having a strong, well-known vetting process and an open forum for the relevant parties from both central offices and the school/unit seeking the new solution is of paramount importance.

It is critical that such an open forum be perceived as open to the best fit alternative, not simply biased towards recommending the lowest cost option, the product offering available through our existing software providers, or the one that addresses the narrowly defined line of business if it does not serve the larger enterprise vision well. This same service function could serve to evaluate whether in-house custom development is the most appropriate course of action in much the same fashion.

The accompanying spreadsheet – Software Selection Requirements – Software Selection Requirements - Sample.xlsx – provides a template that can be modified and used to assess any solution being considered.

Appendix C: Architecture Requirements for System RFPs

NOTE: The following requirements need to be included in any RFP when a new system is being sought, or an existing system is looking to be substantially upgraded. We acknowledge that Northwestern does not yet have all the pieces of this infrastructure in place, and we are fully aware that even 3rd-party vendors who tout themselves as "leaders" may not have these available yet. However, these should be included in the RFP requirements, and the discussions will go from there. Involving NUIT early in the RFP discussion for major systems needs to be standard practice to help avoid misunderstandings between the vendors' claims and the reality of what Northwestern's infrastructure is and what we are expecting them to be able to do.

Northwestern University operates an enterprise system and information architecture that emphasizes flexibility and agility through virtualization and standards-based integration. New software systems introduced into the environment are required to:

a. implement all necessary external interfaces through services with appropriate security (Web Services or REST), and
b. run as one or more virtual machines under the vSphere Enterprise Plus 5.1 hypervisor running in a clustered environment, or run as a software-as-a-service (SaaS).

Supported operating systems are either Red Hat Enterprise Linux version 5.6 and above or Windows 2008R2 Server or Windows 2012 Server.

I. Required external interfaces

The following external interfaces, exposed as SOAP Web Services or REST APIs, are required of the application – either on site or as SaaS:

A. Accounts

Authorizing and Provisioning Accounts

Adding authorized accounts to the application security database must occur through a service. This service will be invoked by an external workflow concerned with management approvals. New accounts should have sufficient identifying information local to the application to assist in troubleshooting, but should not replicate information from authoritative systems unnecessarily. An alternative approach is "first access provisioning", where the person attempts to gain access to the application by presenting an authenticated identifier and the application invokes a Northwestern service to determine if that person should have an account. If the Northwestern service approves, then the application creates the security database entry. However, whenever a new account is created, the application must accept parameters to allow for permissions and information control inheritance such as: (a) to have no permissions to access functions beyond initial login, (b) to be preset permissions to one or more pre-defined security profiles, or (c) to inherit the security profile and ownership of local data files of a previously suspended, named, account.

1. Changing Account Attributes

A service must exist to permit changes to an existing account in a limited number of ways consistent with the business functions and applicable compliance guidelines.

2. Suspending and Restoring Account Access

A service must exist to allow an external system to suspend and restore an account's access to the application.

3. Status of an Account

The application must provide a service which will report the status of an account, including both static information and dynamic characteristics such as storage used, number of access attempts, number of successful access attempts, number of transaction requests for subject data, etc.

4. Removing Accounts

A service must exist to allow an external system to remove an account from the application security database. The service should suspend the account and allow for one of the following treatments of unique data within the account: (a) complete deletion, (b) retention, (c) transfer to another account, and (d) collection into an archive format for later retrieval.

B. Access, Authorization, and Auditing

1. Authentication of credentials is abstracted

The application must invoke a service or standard protocol that abstracts authentication to a Northwestern-provided method. Passwords cannot be held within the application's own security database. Example protocols include Shibboleth, SAML, LDAP, and MS Active Directory.

2. Authorization to application functions

The application must implement an internal "role-based authorization" scheme to allow Northwestern to create logical groupings of application functions that can be approved for an account upon access. These roles would be assigned to the account during either its initial authorization or a subsequent change (by services described above).

3. Auditing of account access

Based upon a global switch and an individual account switch, all access and all business functions invoked by an account must be recorded in a secure logging space. Control functions must be exposed as services and include at least the following primitives:

1. Turn logging on/off
2. Enable/disable logging on a single account
3. Enable/disable logging on all accounts, or all accounts with a given role
4. Enable/disable logging on all accounts with the ability to use a given business function based upon role definitions.
5. Enable/disable logging on all uses of a given business function.

C. Data

1. Updating data

The application must be capable of operating in a real-time integration environment. Data must be handled in a transactional manner, not in a periodic "bulk update" manner. The application must provide services to add, change, or delete data under appropriate security.

2. Initial bulk loading of data

To allow for efficient deployment of the application, there must be a bulk loading utility to initialize the data.

3. Data check-pointing

The application must provide a utility function to copy all database information in a format suitable for future bulk loading as a means to checkpoint, backup, and archive the data whenever desired.

4. Exposing and manipulating data via services

The application must provide a supported method for Northwestern to author services (SOAP or REST) that retrieve, update or delete data from the database. These services would be intended for application-to-application integration (including portal integration). A "supported method" is consistent with the software architecture and eliminates concerns during software update cycles.

II. Portal compatibility

Northwestern operates a portal system for which individual business function "portlets" may be deployed to expose frequently used functions of any application to a navigational dashboard. The application must provide a supported method of participating in this environment smoothly and without concern during software update cycles.

III. Distributed processing environment and DR/BC

Northwestern University has two campuses within the Chicago area that are connected by dark-fiber facilities. Each campus has a data center facility, with the primary facility on the Evanston campus. The University operates multiple 10Gbps links between campuses and data centers.

A. Business continuity requirements

The University uses F5 traffic managers both locally in each data center and globally over the data centers. The University highly values applications that can operate in "active-active" or "active-passive" configurations across the two campuses to allow both site maintenance and protection against a single data center failure. Software systems should either rely upon Northwestern's Oracle RAC infrastructure to replicate data between instances on the two campuses or show how their particular database architecture can support regular database change updates between sites. In the case where the production database site is disabled, the software should be capable of rapid reconfiguration to begin using the replicated instance. It is highly desired that such a reconfiguration be automatic.

B. Business-day maintainability requirements

The University strives to build infrastructure and applications to allow business-day maintenance of physical components, power systems, and cooling systems. The vSphere virtual environment is central to this approach and applications must be prepared to (a) operate in the vSphere environment, and (b) be compatible with "vMotion" to allow transparent movement of the application during production to isolate underlying hardware for maintenance.

C. Single data center redundancy across vSphere clusters

The University uses F5 Local Traffic Managers in each data center to allow applications to spread multiple Web and application servers across vSphere clusters. This approach allows maintenance and upgrade of entire clusters without disruption to application availability. This ability is highly valued in new software applications.

D. Two data center redundancy across vSphere clusters

The University uses F5 Global Traffic Managers within its network to allow applications to spread multiple Web and applications servers across vSphere clusters located in different data centers for protection against certain failures. This ability is highly valued in new software applications.

IV. Data warehouse and reporting compatibility

A. Bulk Data Extraction

In order to facilitate reporting and analytics through a data warehouse, any vendor application should provide open access to its underlying database so that data extraction tools and SQL viewer tools can easily query the entire database tables and extract data in bulk. If the vendor architecture restricts direct access to its database, further analysis should be done to ensure that any bulk data extraction interface provided by the vendor adequately exposes data required for analytics and such an interface is compatible with enterprise ETL tools.

B. Data Model Documentation

Proper documentation of the application data model and a detailed data dictionary with description of each table and field should be provided. The documentation should also include description of how the application layer triggers changes to the underlying data.

C. Customization of Database

The application database should be extensible so that meaningful table names and fields can be added easily to support NU-specific customizations, and so that NU is not restricted to repurposing pre-delivered tables or columns. Using standard nomenclatures will enable easy-to-understand data lineage from the data warehouse to the source system.

D. Change Management - Upgrades and Patches

The vendor should provide adequate release notes and documentation with patches and upgrades to the software such that scope and impact of database changes can be easily understood, and any changes needed to the data warehouse as a result of applying patches and upgrades can be planned in advance of any scheduled upgrades/patches.

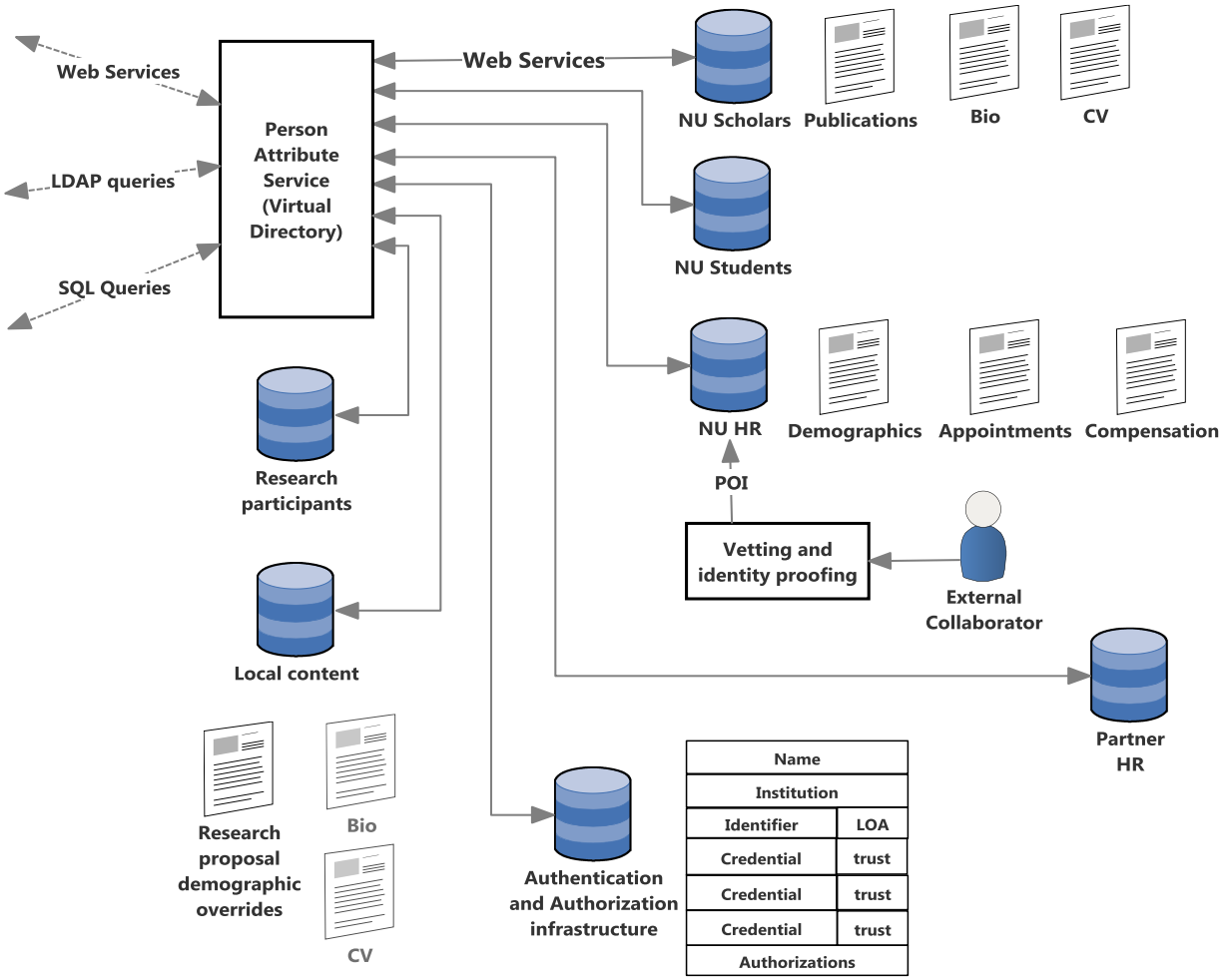
E. Embedded Analytics

A vendor application may also provide embedded analytics and dashboards. Information presented through such delivered reports should be easily customizable to conform to the University's reporting needs. Other portals or applications should be able to consume such data through the web service interface.

Appendix D: Person Attribute Service

The Person Attribute Service would provide information about individuals that was available by agreement from authoritative sources. This repository could be a University-wide service or be limited to just those individuals connected to research, both within the University and collaborators in partner and external institutions. From information provided by the service, applications would make authorization decisions to allow or disallow access to functions or views of data. The information being made available by this service will need to be all inclusive, have the capacity to record multiple affiliations and job titles, and have a complete history of a person's' relationships to Northwestern.

While this service will function much like a repository, in actuality it will be a service platform to virtualize the authoritative sources of information for all individuals and remove any requirement that a requesting application be aware of how the data about individuals was organized or stored in those authorities. The repository would be a "virtual directory" that would accept requests in a number of formats (SQL query, LDAP query, SOAP/REST) and return information about a person on demand.



Concept: Person Attribute Service relationship to authoritative sources

The figure above illustrates the virtual directory concept. A single service point responds to requests about individuals. This programmable platform would define a directory database schema, but the information in the directory would be retrieved in real-time from authorities such as NU FASIS, NU SES, the NMFF HR system, etc. Other sources could be included, such as locally administered content like alternate addresses to use in a formatted CV or other grant paperwork, or study participant profiles.

Virtual directory software is a commercial off-the-shelf product with sophisticated capabilities designed for exactly the purpose intended here. Linkages to authoritative systems would represent the primary development requirement to deploy this approach.

Appendix E: Data Integration

Data Architecture

The current inventory of research administrative systems (see Appendix A on page 26 above) contains a wide variety of transactional systems, some centrally managed, some locally managed, some by third-party vendors, some built in-house. They all generate different information sets to support different business functions.

As shown in the information flow diagrams included in Appendix A, the broad families of data objects flowing between the major systems sometimes originate in one system and then flow to other systems where attributes about the original data may get updated. Those updates may or may not flow back upstream. There are also situations where the same data object may originate in two different systems, and may or may not match.

These fragmentations, duplications, and/or mismatches cause much wasted effort, lead to frustration for people required to enter data yet again or sort out the conflicting values, and makes cross-system reporting difficult.

Adopting a few *best practices in information management* will minimize the risk of inconsistencies in data across multiple systems, and enable better integration to support both business workflows and reporting.

Data Dictionaries/Standards

Standard data definitions ensure consistency in data usage across multiple systems, and reduce risk of misinterpretation when generating reports and using information for decision making. These definitions should be combined into a data dictionary that provides a glossary of terminologies with enterprise definitions of all data elements. If data from one system is transformed into new data elements, the transformation logics and data lineage information should also be captured in the data dictionary.

The information contained in the dictionary is very important meta-data (data about data), and should be visible to everyone in the enterprise. Documentation of data lineage and transformation logic becomes especially important in a data warehouse, where transactional data sets are aggregated or transformed using business logics to enable ease of reporting. It also is important when considering new systems because it shows what the data flows need to be in and out of that system.

Currently, data dictionaries are available for some systems only, and awareness and access to the data dictionaries may be limited. In addition, the types of information captured by the disparate data dictionaries are not standardized. An easily accessible and searchable data dictionary that is vetted by a data governance group would provide a unified repository of terminologies for reliable reference and will ensure appropriate use of data across the research enterprise, both by developers and consumers of information.

Master Data Management

Master data management (MDM) is the practice of defining a single authoritative source for consistently defined business data that is used by multiple business systems. A list of master data in the research administrative sphere would include such items as protocol, proposal, employee/person and area/unit. Currently we have neither practice in place for our research administrative systems – neither a set of

standard definitions and policies for maintaining them, not an architecture that enables the maintenance and usage of these master data elements. Both are required to provide this key foundational unifying element, but developing the definitions and policies for usage is the first step because it's a more complex and complicated process than choosing and implementing a technical solution. (See also Data Governance in the Governance section of the report.)

Managing Data via Real-time Services

Agreeing on standard definitions and usage is a first step in unifying the data into an integrated whole that can be effectively built upon. However, there also needs to be a way within a multi-application environment for that data to be shared. Currently, to the extent that it is done, it is done via batch exports of data that are shipped to other systems for ingestion. This approach introduces time delays, excessive effort to build and maintain these data exchanges, and where there are no batch exchanges, redundant data upkeep, presenting both inefficiency in process and the likelihood that data will become stale and/or out of sync with the authoritative source data.

In order to optimize our research administration portfolio, this needs to change from batch sharing on some time delayed time schedule, to real-time sharing of individual units of data, much of which needs to happen automatically behind the scenes, either when a data element is needed, or when a data element changes.

Integrated Data Warehouse to Enable Reporting and Analytics

Integrating data across the transactional systems is a large and important step. However, it is not a sufficient step to provide the integrated reporting and analytics that are very much in demand.

Data for reporting currently exist in multiple source systems such as InfoEd, SIMS, Maximus, Granite, ACUC, FASIS, and SES. There are even some system-specific data marts in the BI tool that enable enhanced reporting for that particular system. Although all these systems provide some level of reporting, there are still much needed reports and analytical capabilities that require significantly more integration and pre-aggregation than currently exists to ensure simple, intuitive access to various metrics and statuses relating to research enterprise.

Research administrators or dean's offices, for example, need access to all of a PI's research within a department, or they may need an aggregated view of a center or a school. Administrators in the Office for Research may need to monitor certain activities across multiple schools for compliance or risk management, or may need to report on research productivity across a large population of researchers.

An integrated data warehouse repository that unifies the current silos of data for research administration is required to enable cross-system aggregation and integration of data for reporting across the multiple research-administrative systems. Once this unified data warehouse repository is built, it becomes much easier to provision self-service access to the data for reporting and analytics purposes through a BI tool, or to expose data as a service for consumption by applications or portals.

Appendix F: Workflows – Using data to manage the processes of research

A robust services architecture will also play a key role in another critical unifying "application": workflows. (For other discussions of a service architecture and its central role in keeping data integrated, consistent, and available, see also the Data Integration section of Recommended Architecture its corresponding appendix, the preceding Appendix E.)

There are numerous disparate transactional systems in which workflows already exist within the system, presenting transactions to the appropriate party or parties for *approval*. For example, NUFinancials routes purchase requisitions, InfoEd routes proposal applications, and FASIS routes funding requests. In all three cases, a system function exists to post items to a workflow inbox internal to the application and notify the appropriate party via email that the transaction is awaiting approval.

To further unify these processes, these internal processes could be "exposed" and the prompts to take the next step in the process could then be routed by a workflow outside of the system to a unified Inbox on a portal. This workflow would not only populate the message in the inbox, it would also enable the subsequent step in the business process to be taken without having to enter the system from where it came.

To achieve a consolidated inbox on the portal, there is a need to expose these work list items as services, and display them in an application external to their source. Upon clicking a transaction in the consolidated work list, an approval page would be presented to the user (likely the native page upon which approvals are performed in the relevant system) for him/her to take action.

A similar approach might be used to deliver *informational alerts*. For example, such alerts could be used to display a list of approaching protocol end dates, an outstanding effort report, or an over-budget situation on a grant. In these cases, clicking on the message in the combined inbox would launch an application where the relevant data can be viewed (such as a Cognos BI report displaying budget status or an eIRB screen depicting information about the soon-to-expire protocol).

There may be limitations to what we currently have to work with; our Oracle-produced applications can likely meet our vision, however, the workflow/messaging capabilities of the products from our smaller vendors should be reviewed. Some applications may not have native workflow capabilities to present items in an inbox in their native applications and/or notify the relevant individual that such items are awaiting action. In such cases, services would need to be built to surface the need for action, and additional services would need to be built to populate the consolidated inbox. Similarly, some of the currently deployed enterprise applications might not have their workflows built with open standards, and other workarounds might have to be deployed – e.g. gathering data for the Inbox alert from data marts – that would limit our ability to make inbox items real-time in nature.

Looking ahead, the ability to perform integrations in this manner should be a core requirement in our software selection methodology, and, in order for these processes to be seamless for the user, an integrated single sign-on environment is essential.

Workflow Use Case: Research Proposal Creation

A key improvement planned for the administration of research is to simplify and organize the entry points for creating new research proposals. The goal is to reduce the effort required by researchers to

describe, compile, and gain institutional backing for a proposal to be submitted, and to make accurate and up-to-date information about the proposal and research available at all times. This section describes a fully automated online version of this business process.

The fundamental tool for this improvement is electronic workflows at multiple levels of the administrative process. In an online workflow, the pathway of steps is dictated by the context of the researcher, the subject matter, and the target agency just as it is in a human-driven workflow; however, the researcher is not required to discover and navigate the unique pathway resulting from those parameters. Instead it is codified into a top-level workflow, and using templates, forms, and workflow logic, this tool acts as an "expert system" to guide and serve the researcher.

Figure 3 illustrates how layered workflows can organize the work of composing and submitting a research grant.

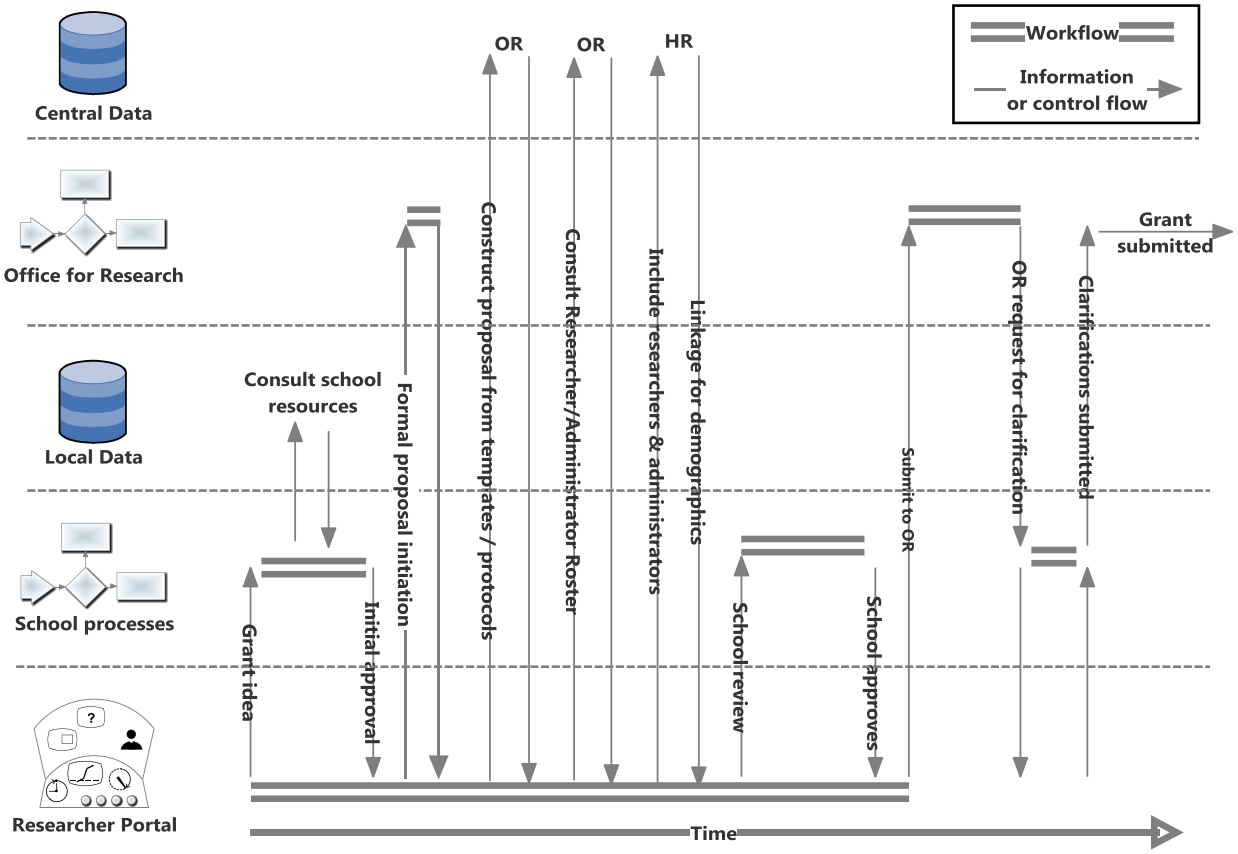


Figure 3 - Conceptual approach: multi-level workflows to organize work

Within the portal environment, the lead researcher enters a workflow spanning phases from initial concept to submission of the finished proposal for approval. This workflow is automatically customized to the researcher's affiliations with school(s) and the particular type of grant proposal (through a template provided by Office for Research (OR)). The researcher does not need special knowledge or

assistance from the school or OR to navigate the process – it is codified into the workflow itself. See Figure 4.

At appropriate steps, the workflow itself triggers other workflows – certainly in the school dean's office, the Office for Research, and perhaps in other institutions – to follow approval pathways laid out previously and relevant to the context of the grant. See Figure 4 below.

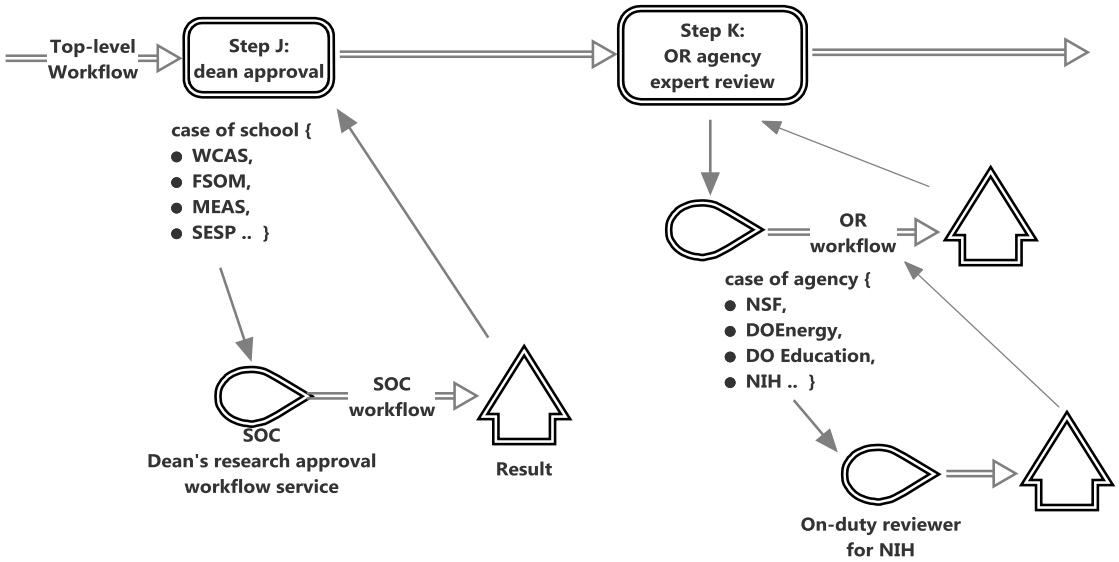


Figure 4 - Context-driven workflows

Via the researcher's consolidated Inbox on the portal, s/he is made aware of these approvals as they happen, and may monitor the progress of the process.

The process of proposal creation can be further enhanced if information required for the grant formulation is available from a mixture of central and local databases via services which can be invoked by the workflow. Rosters of co-investigators, key staff, service centers, etc., become point-and-click additions based upon context and the researcher's goals.

This package of selections travels with the grant proposal through the workflow and eventually into post-award provisioning of services, budgets, and triggering of additional workflows to complete administrative tasks prior to the start of research proper. See Figure 5 below.

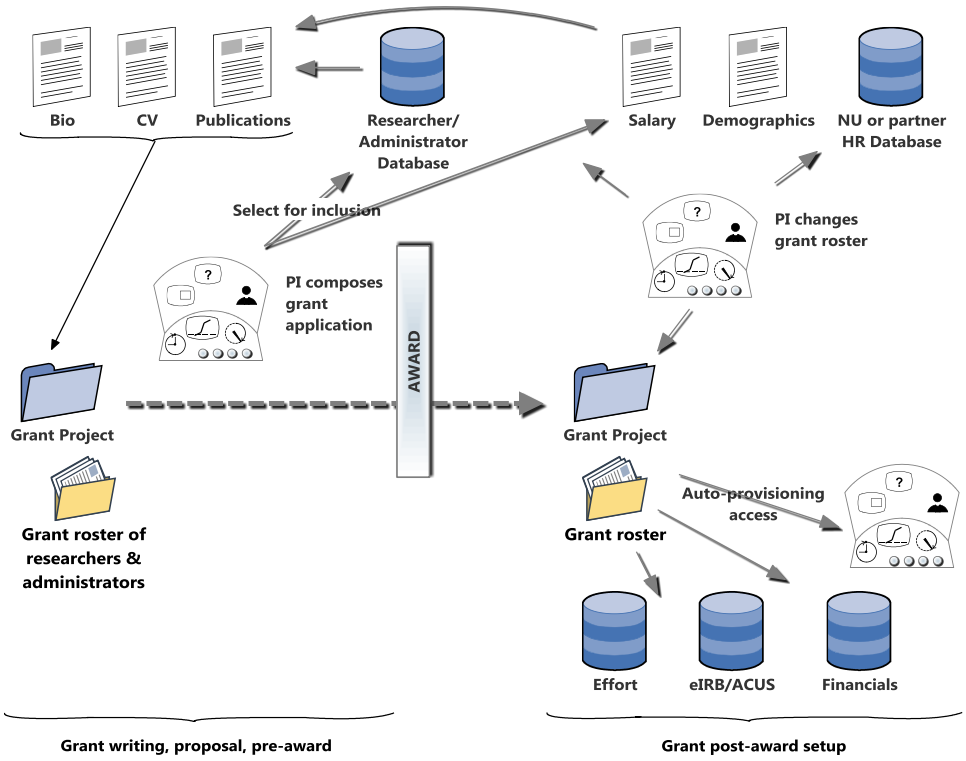


Figure 5 - Conceptual approach: managing participation in research administration

The exceptional aspects of this workflow-based solution result from the flexibility of a software solution, the multi-level approach, and ability to instrument processes and measure performance:

- Multi-tier workflows recognize that the research process involves distributed innovation, followed by increasingly central reviews and approvals, award, provisioning of central services, distribution of control to the researcher during the life of the grant, and eventual close-out. The means of review may vary between schools or within disciplines – reflecting unique workflows. See Figure 6 below.
- Codifying how grants are composed, approval gates, and the guided assembly of the correct materials allow reliable replication of the correct approach to given subject matter, requirements of individual agencies, and necessary materials to support human or animal subjects.
- Electronic workflows can be modified easily and new ones can be deployed transparently to the research community. A new review requirement for research involving a particular material can be inserted at the appropriate point – and it will be immediately effective without wide-spread training.
- Viewing workflow statuses from a portal dashboard will efficiently inform any authorized person of the standing of the grant over the course of its lifecycle.
- Materials assembled during the composition of the grant can seamlessly flow to new workflows initiated after award.

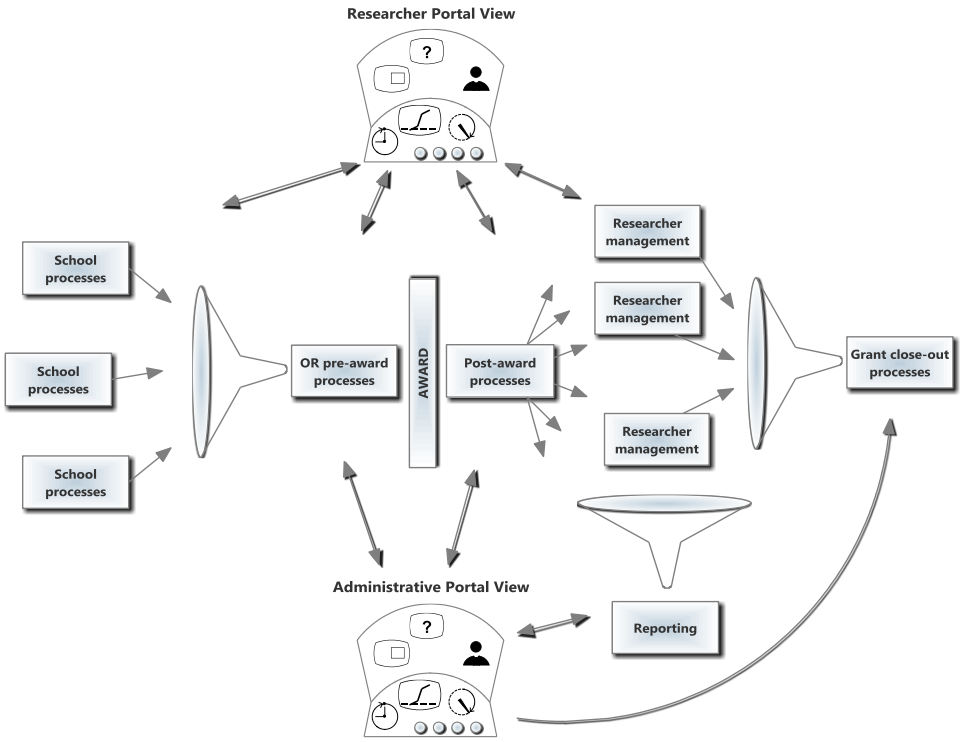


Figure 6 – Conceptual approach: the research administration process

Appendix H: Unifying Views – Portal, Reports/Analytics

As the number of different systems, and their corresponding sets of data, grows, the importance of having unified views into these systems and their data also grows. Each of the topics covered in this section – portal, reports and analytics – provide different methods of organization that create value by calming the chaos into patterns and pathways.

Portal

The difficulties imposed by the breadth and lack of organization of the research administrative process at Northwestern is a familiar refrain, and the unifying views that portals can provide would offer great value for everyone concerned with the administration of research: PIs, the PI's research team, research administrators, members of the Office of Research, Deans.

In order for a portal to succeed in serving as an enabling home base, it needs to organize the chaos of multiple systems in a personalized manner, provide an entry way into a critical mass of systems, resources, and information, and expose information about key processes in which that particular user is involved.

The sections below describe some of the features that a portal, deployed in the environment described in this report, can provide to facilitate the process of research. In all of these cases, the solution being proposed depends on the architecture described heretofore, e.g. a robust service architecture, single sign on, data integration, robust but flexible security mechanisms, and ease of use considerations such as browser interoperability and support for access from mobile devices.

Providing a Critical Mass

Because the success of portals is so much tied to providing a "critical mass" of features, and because researchers and research administrators also have other roles at the University, the portal for research administration should be part of the overarching University portal.

Personalization

One size does not fit all when it comes to portals. The more the portal is personalized for each individual user, the better its chances of being successful. Personalization comes on two main levels: organization of resources, and access to content. In the tools mentioned below (e.g. in an alert Inbox), the content that will be displayed will only be for that person, but there are several basic ways content could be provided to give a personalized view on resources and content:

- The array of resources displayed to the user could be based on their role within research (PI, research administrator, etc.) and/or the type of research they do (clinical, base science, etc.).
- Taking this one step further, many of the choices about how these resources are arranged on the portal pages, or which ones are included at all, could be left to each user. There will be some parts of the portal that will not be removable, and some that will probably not even be movable, and studies have shown over and over again that most users take the default view of any software interface regardless of their ability to change it. Nevertheless, the ability to customize portals is a standard part of this software these days, the point is to make the site as user-friendly as possible to incent usage and some people will take advantage of these options., in general, higher levels of personalization are better than lower levels.

Unified view (dashboards)

As it relates to information delivery, dashboards and simple reports should be included in the portal, tailored to the person who is using the portal. They should be easy to interpret at a glance, and quickly inform the researcher about an important aspect of the status of his/her current and proposed research.

Alert Inbox

In order to organize and facilitate the many processes any person involved in research must traverse, each person should have their own Inbox on their portal page where s/he can view alerts and take action on items awaiting approval or follow up. Examples are manifold: approving financial activities, receiving alerts before a grant account is overspent or that an item has been received, monitoring protocol statuses or training statuses, being notified an effort report or sponsor report due date is approaching, or being told what the next step is in submitting a proposal. Providing this Inbox in the portal, and "exposing" pending workflow steps or informational alerts in the same place, prevents people from having to go to each individual system to do their tasks, eliminates clutter in email Inboxes, and helps to reduce risk of lateness.

Workflow Wizards

Wizards or other guides should provide intuitive starting points to conduct common transactions, whether initiating a new proposal or study, purchasing materials, or fulfilling compliance obligations throughout the lifecycle. While it would likely prove infeasible (either due to effort required or technical feasibility) to "front-end" our enterprise applications in their entirety, the creation of "wizards" to aid users in the performance of common (and often complex) business processes is likely to drive increased adoption of the systems by researchers and simplify the environment for all users.

These wizards can serve to illustrate the business process in a more graphic nature, indicating to a user not only where to begin, but also that there are Y steps involved and they are currently on step X of Y. A wizard-style interface will also serve to bridge the multiple systems that are often required to be used in tandem to complete a business process.

These wizards will likely require custom development or development external to the native enterprise systems. This custom development layer will also offer a prime location in which to embed business rules that can both determine which steps must be completed based on data provided by the user and enforce edits that can keep data in sync across systems.

An underlying web services architecture will again be necessary either because the wizard interface will need to navigate deeply into one or multiple native enterprise systems in order to perform a function, or, where it is technically feasible and deemed acceptable within the data maintenance rules of that system, data may be sent to the native application without requiring the user to visit its front-end interface.

Using wizards also brings with it the opportunity to include content not housed elsewhere in our enterprise systems environment. Examples might include using web forms to capture additional data, the ability to populate and/or attach related documents and forms, and the ability to nest training and other instructional materials in the natural flow of the business process.

Reports and Analytics

There are numerous opportunities to serve up content targeted towards researchers, their teams, and their administrative staff throughout the lifecycle of a research grant/contract. Dashboards and simple reports should be presented at the portal layer; additional elements should be embedded into the transactional systems where relevant; more in-depth reporting should be available in a single click from either of these sources.

The reports for research administration should serve six distinct types of users:

1. the primary investigator conducting or participating in research, who may ask questions such as "How much money do I have left to spend on a grant account?" or "How many graduate students do I support?";
2. the business administrator supporting one or more primary investigators within a school or a department, who needs to monitor the status of grants and proposals for financial or research compliance;
3. an administrator such as a department chair who might be interested in viewing all or a subset of research activities within a department for tracking commitments such as those related to cost sharing, effort, or space;
4. a dean's office staff interested in various administrative and operational aspects of research administration across all departments - both within the school and outside of school (typical of interdisciplinary research) - to make decisions around allocation of scarce resources such as space or retention of faculty;
5. central administration staff, for example, in the Office for Sponsored Research to monitor and manage a smooth research operation while ensuring compliance;
6. senior administrators in the central offices such as the Budget Office or Office for Research to support strategic decision making who may be looking for more aggregated information.

All of these features depend on the ability to integrate data across diverse sources and prepare it for presentation in formats that are intuitive to the researcher. Use of common identifiers throughout the underlying data repositories and the deployment of a security model that grants access to data throughout the full lifecycle of the grant/contract will be critical elements to enabling these functions.

In order to enable easy reporting with consistent data for all the various user constituents, the underlying data warehouse structure should be integrated and structured in a design that supports quick access to both aggregated and detail transaction view of the data while maintaining consistency in data output for the various levels of aggregations. A powerful BI tool such as Cognos (the enterprise standard at NU) will provide intuitive navigation and user defined reporting capabilities to access the data warehouse. Data security becomes a key component of the architecture, where different user groups need access to different subsets of data. A data access policy based on roles would greatly enhance the implementation of security rules.

In addition to simple reporting, other functionalities such as alerts (email based or dashboard notification) built in the BI tool and forecasting capabilities (for example, "What if I paid for two graduate assistants?") powered by existing tools such as TM1 will provide additional analytics capabilities users desire.