

arXiv NG Phase I Project Charter

PROJECT DIRECTOR	Oya Rieger
PROJECT COORDINATOR	Gail Steinhart
START DATE	12/1/2016
END DATE	6/30/2018
APPROVED BY	arXiv-NG Steering Committee: Oya Rieger, Sandy Payette, Gail Steinhart, Martin Köhler (MAB), Dave Morrison (SAB)
APPROVAL DATE	3/21/2017
APPROVED VERSION	1
CURRENT VERSION	2
LAST UPDATED	4/23/2018

arXiv is an openly accessible, moderated repository for scholarly articles in specific scientific disciplines. arXiv provides rapid and perpetual access to content, without any fees and with support for users around the world.

DESCRIPTION

Purpose: The overall goal of this project is to renew arXiv's technical infrastructure, in order to continue to fulfill arXiv's mission of providing rapid dissemination of research findings at no cost to readers and submitters. Work will proceed in two phases; the purpose of Phase I (18-month) is to develop a complete plan to renew arXiv's technical infrastructure, and to deliver either a working proof-of-concept next generation system (arXiv-NG), or selected production-ready modules for a next generation system. Phase II will see through to completion the work initiated in Phase I, resulting in a fully functional, production system.

Rationale: While the current arXiv system has proven remarkably reliable even as the number of users and submissions has continued to grow, it has also become very difficult to extend and modify the system, and requires staff with knowledge of programming languages that are becoming obsolete.

Anticipated results: An arXiv infrastructure that will be less work to maintain and utilizes modern, standard programming practices, resulting in greater capacity to implement improvements and new features, and support the development of new features by others. We will retain the essential features users expect of the current system, while introducing improvements with minimal disruption to the user experience. The finished system will be user-focused, sustainable, and production-ready. The end result will be a superior user experience for readers, submitters, moderators, administrators, and arXiv member-supporters.

The activities of Phase I are making decisions regarding overall approach to the rebuild, conducting rigorous evaluations of candidate technologies, developing a design for a renewed system, producing working components or a working proof-of-concept system, developing a new five-year business and governance plan with capacity to support the new system, and writing successful grant proposals to complete the remainder (Phase II) of the project.

Approach: We expect the complete, multi-phase design and development of a next-generation arXiv (arXiv-NG) to take approximately 3 years, with additional funding and work required to deliver a complete system. Phase I work (technology assessment, comprehensive planning, and early implementation) is essential to inform subsequent funding proposals.

As a result of consultation with the arXiv-NG IT advisory group and arXiv stakeholders, we favor a renewal strategy that includes a mix of in-house development and use of off-the-shelf services and technologies. Still to be determined is whether our process should be a complete rebuild and migration, or a stepwise renewal of the existing system. A thoughtful approach to these decisions, and technology evaluation and selection, dictate that we adopt an iterative approach to our testing, design, and planning activities in Phase I.

Routine maintenance and modest improvements to the existing arXiv system will continue throughout Phase I.

Alternatives considered: The alternative to rebuilding arXiv's infrastructure is simply not to. This is not a viable option. With current staff and financial support levels, it is barely possible to keep up with the most urgent maintenance and improvements, let alone address new improvements and feature requests.

DELIVERABLES AND MILESTONES

1. Formal project charter – purpose, approach, process.
2. Technical requirements and technology selection and justification.
3. Technical architecture diagram and first principles.
4. Working and documented proof-of-concept next generation system (arXiv-NG), or selected production-ready modules.

5. Report on partnership opportunities and tangible progress in establishing such alliances.
6. User testing strategies and results from early implementation experiments.
7. Comprehensive project plan (through Phase II) to outline the objectives and timelines to transition the current arXiv service into arXiv-NG.
8. Documented review of workflow, policy, and moderation.
9. Communication strategies for engaging and informing key stakeholders.
10. Draft business and governance models.
11. Grant Proposal(s) for next phases.

EXCLUSIONS

The following are not explicit objectives of Phase I of this project:

- Expansion of arXiv into new subject areas. Subject area expansion is initiated by the arXiv Scientific Advisory Board (SAB), and executed by the current arXiv IT and operations teams.
- Development of standalone preservation infrastructure for arXiv. While we aim to develop robust, “preservation friendly” infrastructure, the preservation function is external to arXiv’s core infrastructure.
- A fully functional production system. This is a goal of Phase II.
- Research on the arXiv corpus.
- Addition of new features. The project’s main focus is on sustaining support for and improving current features; new features will be implemented very selectively.
- Reconsidering arXiv’s moderation practices and policies. While new tools will be developed to make moderation work easier and more efficient, reconsideration of policies and practices is out of scope in this context. The SAB, Scientific Director, and Operations Manager may pursue these issues separately from arXiv-NG development.

GOVERNANCE, ROLES and RESPONSIBILITIES

Oya Rieger, arXiv Program Director and Principal Investigator, has overall responsibility for this project and is responsible for business planning efforts. The arXiv-NG Steering Committee has decision-making authority. More complete information on roles and responsibilities is summarized below.

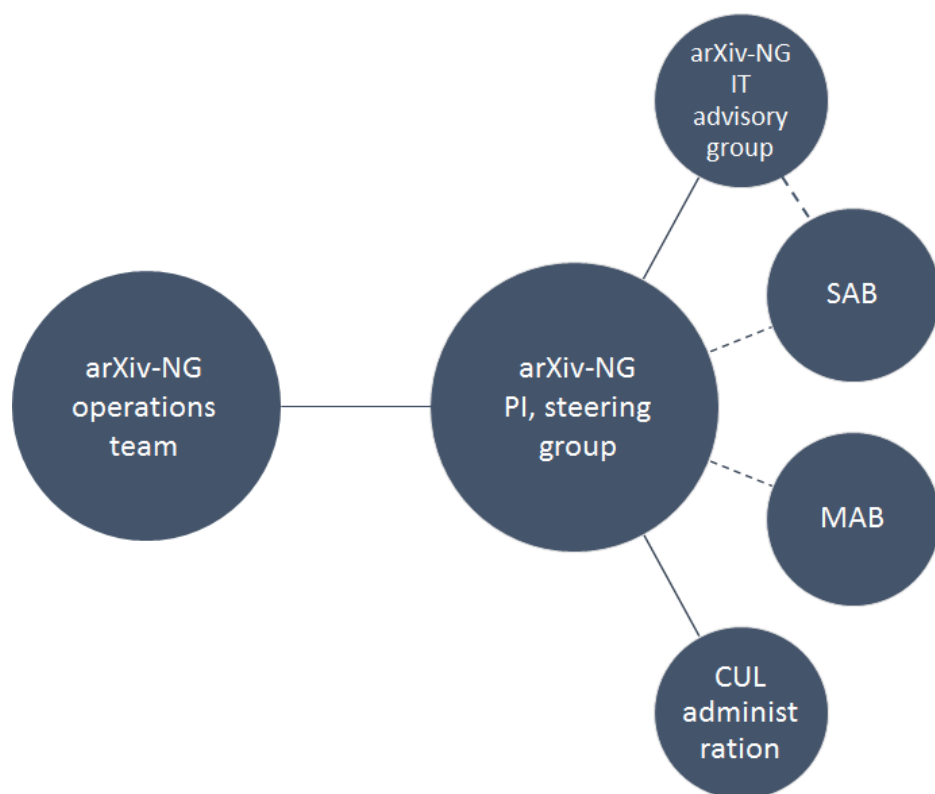
arXiv-NG groups and teams	Roles and responsibilities
Program Director and Principal Investigator: <ul style="list-style-type: none"> • Oya Rieger 	<ul style="list-style-type: none"> • Has overall responsibility for project. • Leads business planning efforts. • Has final decision making authority in consultation with the Steering Group (shared responsibility with Scientific Director, currently vacant).
Steering group members: <ul style="list-style-type: none"> • Oya Rieger, PI • Sandy Payette, CTO • Gail Steinhart, Project coordinator • Scientific Director (vacant) • Dave Morrison, SAB • Martin Köhler, MAB 	<ul style="list-style-type: none"> • Provides strategic advice to PI especially in regard to decision-making on high-level issues. • Counsels on reconciling different stakeholder perspectives. • Contributes to developing communication strategies with various stakeholders. • Represents interests of relevant stakeholder groups.
arXiv-NG Operations team: <ul style="list-style-type: none"> • Sandy Payette, CTO • Martin Lessmeister, Lead developer • Gail Steinhart, Project coordinator • Jim Entwod, Operations Manager • arXiv-NG developers 	<ul style="list-style-type: none"> • Implements and carries out project. • Forms appropriate partnerships. • Takes direction from and communicates regularly with arXiv-NG steering group. • Communicates regularly with the current arXiv operations and IT teams, fostering understanding, support and application of arXiv-NG’s “first principles.” • With other teams, seeks additional funding for next phase. • Identifies and relays problems, impediments, risks in a timely manner to PI and SG.

arXiv-NG IT advisory group: <ul style="list-style-type: none"> • Alberto Accomazzi, NASA ADS • Paul Ginsparg, Cornell, arXiv founder • Robert Hanisch, NIST • Dave Lifka, Cornell University CIO • Mark Matienzo, Stanford University Library • Matthew McGrattan, Digirati • Thorsten Schwander, SLAC /INSPIRE 	<ul style="list-style-type: none"> • Provides input on technology and partnership choices.
Gerald Beasley, Cornell University Librarian	<ul style="list-style-type: none"> • Has overall responsibility for arXiv's obligations. • Provides institutional support and resources for arXiv (HR, business services, legal, etc.). • Is final arbiter for arXiv and arXiv-NG decisions.

Additional stakeholder groups

- Scientific Advisory Board (SAB): The SAB provides input as requested via representatives to NG steering committee, NG IT advisory group, and PI.
- Member Advisory Board (MAB): The MAB provides input as requested by NG steering group and/or the NG operations team.

Relationships between teams and stakeholder groups are depicted below. Solid lines indicate formal relationships between arXiv-NG groups and teams (PI, steering, operations, IT advisory, CUL administration), dotted lines indicate additional lines of communication to other stakeholder groups (SAB, MAB).



COMMUNICATION

A separate communication plan will be prepared.

RESOURCE REQUIREMENTS

Staff resources

TEAM ROLE	NAME	SUPPORT
Project Director	Oya Rieger	CUL supported
Project Coordinator	Gail Steinhart	Grant & CUL supported
CTO	Sandy Payette	CUL supported
arXiv lead developer	Martin Lessmeister	Grant supported
Lead system architect	Erick Peirson	Grant supported
Developer	Jaimie Murdock	Grant supported
Operations Manager	Jim Entwood	CUL supported
Scientific Director	Steinn Sigurdsson	CUL supported

RISKS AND CONSTRAINTS

A separate risk analysis will be conducted.

ACCEPTANCE CRITERIA

Completion of all deliverables, including a proposal to support Phase II.

CHANGE LOG FOR THIS DOCUMENT

- 4/23/2018: updated staffing information

APPENDIX. Overview of project phases

We anticipate a complete, multi-phase design and development of a next-generation arXiv

(arXiv-NG) will require approximately 3 years. This project charter applies primarily to the initial 18-month Phase I (12/1/2016-5/31/2018) for technology assessment, comprehensive planning, and early implementation of a new infrastructure for arXiv. This early work is essential to inform subsequent funding proposals to complete the second phase (also 18 months) of development and testing, and a smooth and complete transition to a next generation arXiv. We are already actively working on the fundraising front as we realize that the overall success of this project will be defined by the successful completion of both phases.

It is our intention to continue uninterrupted operation of the current arXiv system, and essential, modest improvements will continue to be made. We have yet to decide whether our rebuild strategy will be to a) isolate components of the current ("classic") arXiv system, build new equivalents, and carefully integrate the new components into the existing system, or b) build an entirely new system and migrate content in the classic system to the new one upon completion and successful testing. An approximate timeline for the concurrent maintenance of the classic system and development of arXiv-NG follows. The uncertainty (?s) indicated in the timeline has to do with the choice of the a) or b) scenario - a) implies continuous operation of the classic system while improvements are made to it, while b) implies a shut-off and migration date.

[illegible]

Project charter	X											
Communication plan	X											
Technical requirements and principles	X											
Workflow diagrams and mapping	X											
Technology evaluation	X	X										
Partnership evaluation	X	X										
Selection of framework and technology	X											
Technical architecture - design, diagrams, APIs			X									
Comprehensive transition plan			X									
Draft business and governance models	X	X	X									
Grant proposal(s) for Phase II		X	X									
Usability testing plan			X									
APIs/UIs/prototypes for submission, moderation, access	X	X	X									
APIs/stubs for core modules - storage, db, registries, index	X	X										
APIs/integration tests for search, TeX, other services		X	X									
User testing	X		X		X	X		X		X		
Iterative development I	X	X	X	X	X							
Iterative development II						X	X					
Iterative development III								X	X			
Complete transition										X	X	

* Each project period represents three months.