Progress Report:
Building Interactive Digital Libraries of Formal Algorithmic Knowledge
– Project Overview –

Robert L. Constable – Cornell

James Caldwell – Wyoming

Jason Hickey – Cal Tech

Outline

- The DoD URI BAA
- Our Proposal
  - Goals
  - Strategy
  - Benefits
- Challenges
- Reply to challenges
- Progress in Year One
  - Cornell
  - Cal Tech
  - Wyoming

Department of Defense University Research Initiative

ONR BAA – Digital Libraries for Constructive Mathematical Knowledge

"Knowledge about algorithms is derived from discoveries in the mathematical sciences that can be expressed constructively. Software assurance depends on mathematical understanding and experience in constructive mathematics and its expression."

"Systematically developing an infrastructure for this knowledge of algorithms – as in a digital library – will contribute to higher quality software and greater confidence in program construction. This development begins by codifying constructive mathematical knowledge and engaging the research community in pursuit of this national infrastructure."

Research Concentration Areas

1. Proof checking of the standard body of computational mathematics
2. Catalog principal math concepts used in computing
3. Investigate a suitable base language and logic
4. Provide library support for routine aspects of formalization
5. Investigate assured interoperation
6. Investigate reflection
7. Study issues of consistency and maintenance among library collections
8. Investigate appropriate user interfaces
9. Explore innovative correctness metaphors
10. Consider economic issues (as in publishing)

Characteristics of BAA

The research concentration areas have three aspects:

- Building infrastructure for a formal library of computational mathematics (2, 3, 4, 5, 7, 8, 9)
- Creating formal content (1, 3, 6)
- Applying formal content (preamble, 9, 10)

What Does “Formal” Mean?

The BAA refers to machine-checked mathematics presented in a consistent formal logical theory that is implemented.

This meaning of “formal” is technical. It is more narrow than what many people mean in daily use.

Objective:

To create a digital library of algorithms and constructive mathematics useable for program and software construction.

Our Proposal and Project

"Building Interactive Digital Libraries of Formal Algorithmic Knowledge"

Goals

1. Build a semantics-based interactive logical library infrastructure
2. Create, collect and organize formal computational mathematics content
3. Apply the formal interactive DL in designing and creating reliable software (especially for CIP/SW)

Strategy

1. Attract a community of contributors who share formal knowledge and the connected mathematically literate articles
2. Account for correctness in a multi-logic, multi-prover (including tactic-style) environment
3. Provide semantics-based library services at many scales

Challenges and Problems

1. Community using formal proofs is relatively small
   - Market for formal proofs is small
     – proof technology not widely used in software
     – proof technology not widely used in science and math
     – proof technology not widely used in education
   - Formal proving is still hard work
     – expansion factor
     – shallow base of basic mathematical facts
     – demanding skill set (programming + math + design)

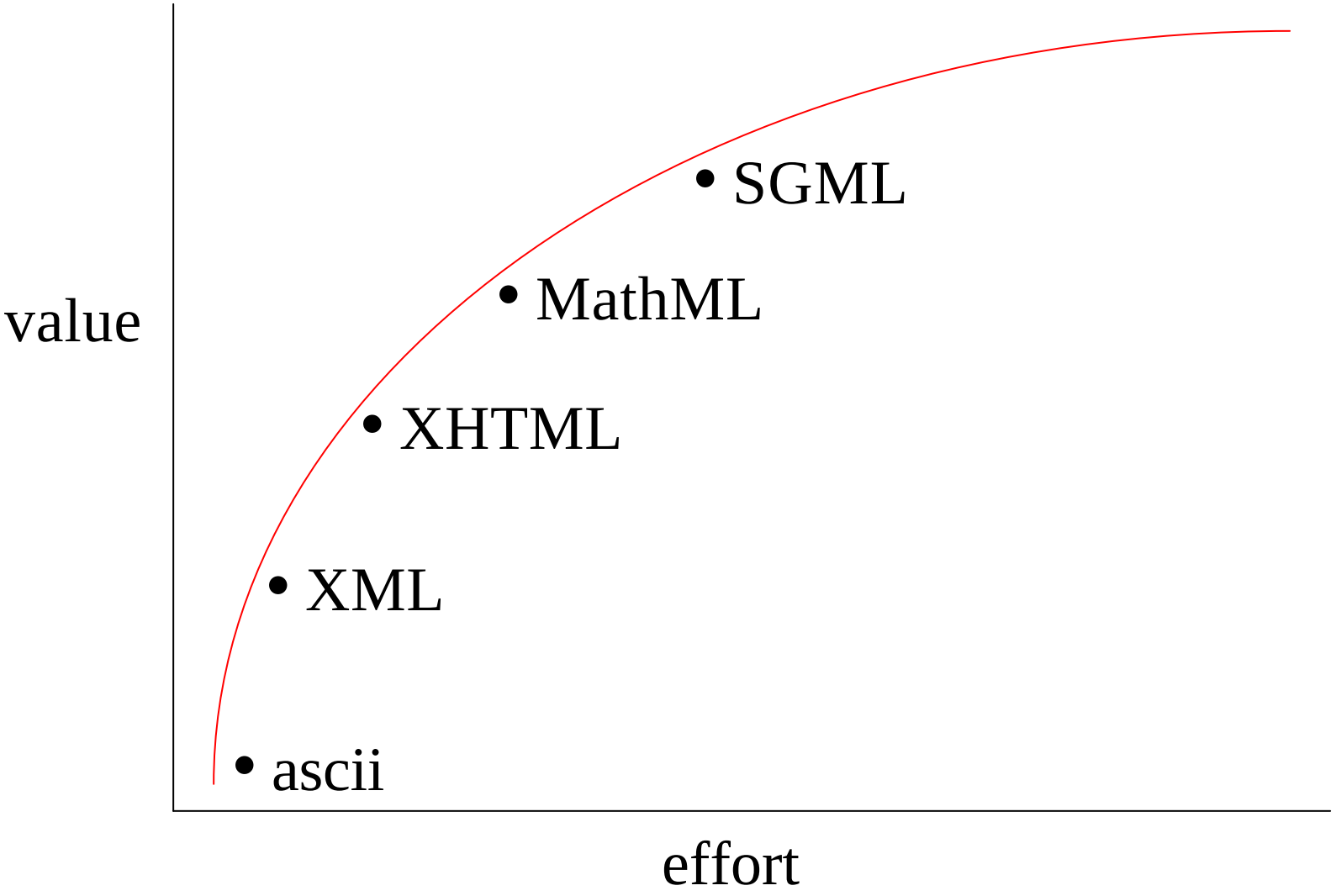
Challenges and Problems

2. Community is disconnected
    - Each group uses a different system
    - Almost no sharing (logical difficulties, practical ones)
    - Systems change or go extinct

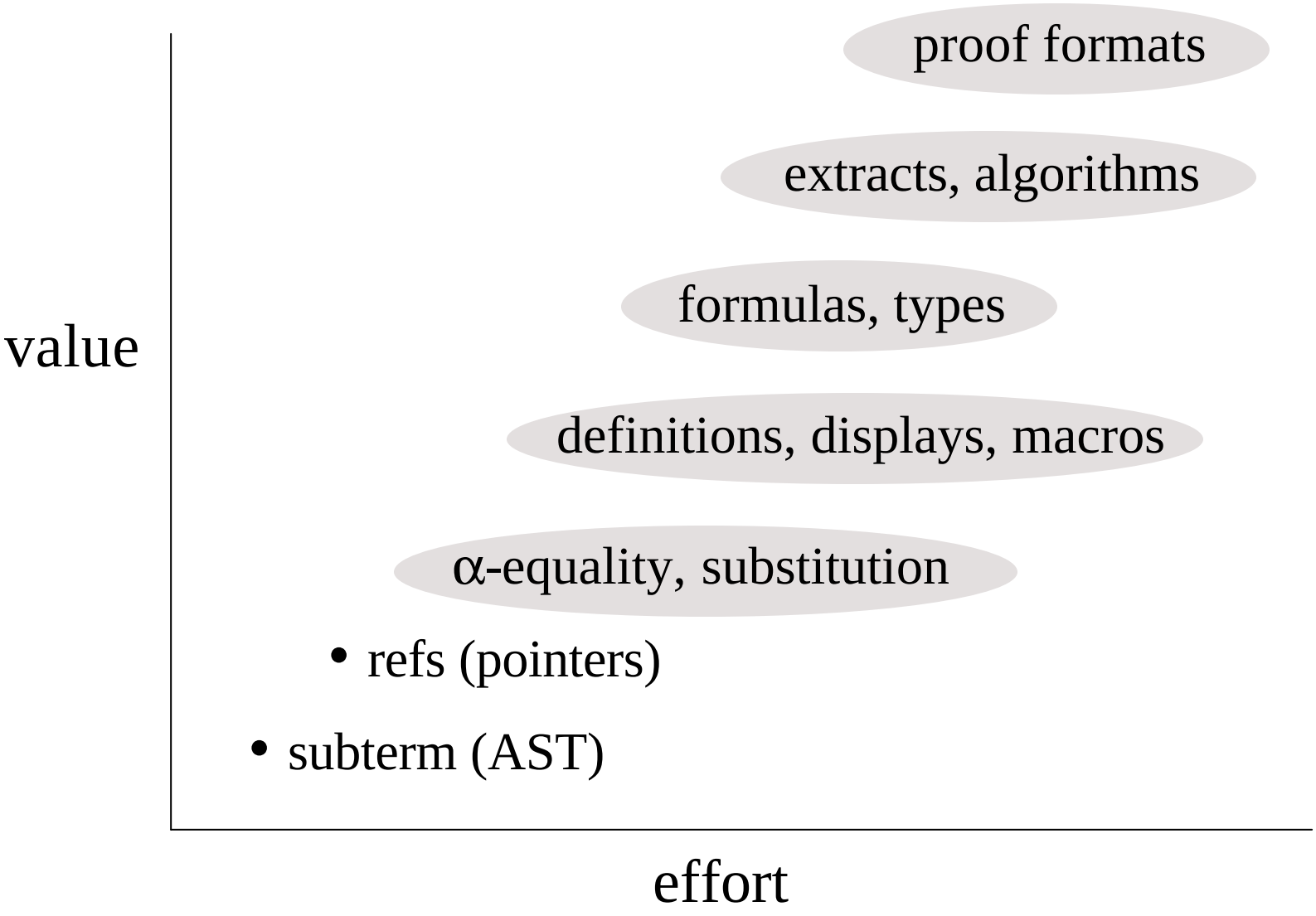
Digital Library Approach to the Challenges

1. Widen the community by
   - library will increase the services provided
   - library will decrease the effort to create proofs (seen from experience)
2. Connect the community through a common service – the digital libraries approach

DL Shared Data Formats



Formal DL Data Formats



Terms (Abstract Syntax Trees)

$t = op(t;\cdots;t)$ for t a term

$Term = op \times Term\ List$

with binding structure

$op(\bar{v}_1.t_1;\cdots;\bar{v}_n.t_n)$ $\bar{v}_i$ list of binding variables

$Op = OpName\{i_1 : F_1;\cdots;i_k : F_k\}$

i can be reference objects or values

Progress in Year One

- Cornell
  - conceptual basis for FDL
  - prototype Formal Digital Library
  - targeted content creation (Nuprl, MetaPRL)
  - applications to DoD problems
- Cal Tech
  - content production in MetaPRL in CZF
  - MetaPRL development
  - compiler in MetaPRL
- Wyoming
  - content production – graph theory
  - primitive proof checker in ACL2

Characteristics of Year One Progress

- Prototype FDL has gone farther than promised
  - includes PVS
  - has a customer (ORA)
  - Nuprl and MetaPRL are only clients
  - extensive advanced design documents
  - reference prototype is small
- Content development has been cooperative
  - strong ties through MetaPRL
  - JProver has customers

Benefits to Society

- Basis for highly reliable and responsive software
- Acceleration of scientific discovery
  - mathematics
  - computer science
  - computational science
  - meta mathematics
- Wider access to content (participatory science)
- Topics in a new science of information
  - formalized mathematics publication
  - scholarly publication in general (arXiv)
  - quantitative meta mathematics

Questions from Dr. Ralph Wachter

1. What will the library look like? How will it be used?
2. How can users make contributions to the DL?
3. How can user interaction with the DL be facilitated?
4. On what criteria would users be able to search?
5. How will the DL be maintained?
6. What will the internal and external forms of DL artifacts look like?
7. What about the issues of intellectual property rights in the DL?
8. How will the library be used in CIP/SW?

What will the library look like?
How will it be used?

View as a reader

View as a formal contributor

View as an article writer

View as a connected process

Mathematical essence:

$$D \rightarrow Term(D)$$

basic operations

How will users search?

Search mechanisms are not built in, by design.

They are part of content.

There will be many search mechanisms, e.g.

- indexes
- tables of contents
- clustering algorithms
- word-based searches
- etc.

How will correctness be maintained?

The library does not guarantee correctness.

It provides accounting mechanisms.

There will be sound theories and there might be experimental combinations of subtheories or individual results.