The way forward is…

ONLINE

The way forward is…

PURPOSEFUL

The way forward is…

REUSABLE

The way forward is…

ENGAGING

Does the BSC *WANT* to be a player in this space, and to what extent *CAN* it play?

NAMES ARE THE KEY TO THE BIG NEW BIOLOGY: A BOTTOM-UP SOLUTION TO CREATING A MACROSCOPE

David P. Shorthouse (dshorthouse@mbl.edu)

Dmitry Mozzherin

David J. Patterson

The Marine Biological Laboratory
Center for Library & Informatics

Cell PRESS

Opinion

Names are key to the big new biology

D.J. Patterson[1], J. Cooper[2], P.M. Kirk[3], R.L. Pyle[4] and D.P. Remsen[5]

[1] Biodiversity Informatics, Marine Biological Laboratory, Woods Hole, MA 02543, USA
[2] Landcare Research, PO Box 40, Lincoln 7640, New Zealand
[3] CABI UK, Bakeham Lane, Egham, Surrey, TW20 9TY, UK
[4] Department of Natural Sciences, Bishop Museum, 1525 Bernice St., Honolulu, HI 96817, USA
[5] GBIF, Universitetsparken 15, Copenhagen Ø, DK 2100, Denmark

Those who seek answers to big, broad questions about biology, especially questions emphasizing the organism (taxonomy, evolution and ecology), will soon benefit from an emerging names-based infrastructure. It will draw on the almost universal association of organism names with biological information to index and interconnect information distributed across the Internet. The result will be a virtual data commons, expanding as further data are shared, allowing biology to become more of a 'big science'. Informatics devices will exploit this 'big new biology', revitalizing comparative biology with a broad perspective to reveal previously inaccessible trends and discontinuities, so helping us to reveal unfamiliar biological truths. Here, we review the first components of this freely available, participatory and semantic Global Names Architecture.

The value of taxonomy to a biology that is changing

'New Biology' is a vision [1] of a discipline evolving to become considerably more data-intensive as it accommodates increasing amounts of under-analyzed data from wide axis by which all biological information might be organized [5,6].

Taxonomy is supported by 5000–10 000 professional taxonomists worldwide, http://www.gti-kontaktstelle.de/taxonomy_E.html [7]. This 'team' [8] is united by principles founded in the codes of nomenclature. Taxonomists discover and describe biodiversity, arrange species into classifications with sensitivity to phylogenetic insights, are aware of all of the literature that bears on the identity of the taxa, and provide services to those who rely on authoritative information. However, many taxonomists feel unable to meet the expectations of the discipline, home institutions or exasperated users [9,10], and even believe that taxonomy as a scientific discipline is in danger of extinction [5,11,12]. Others argue that the 'information age' offers new opportunities to serve those who depend on taxonomic knowledge [6,13–15], and that using taxonomy to manage on-line biological information can reinvigorate the discipline [16]. A small community of innovative taxonomists, computer scientists and data managers (collectively 'biodiversity informaticians') are pursuing this vision and are

TREE 25(12): 686-691
doi:10.1016/j.tree.2010.09.004

David J. Patterson

Dmitry Mozzherin

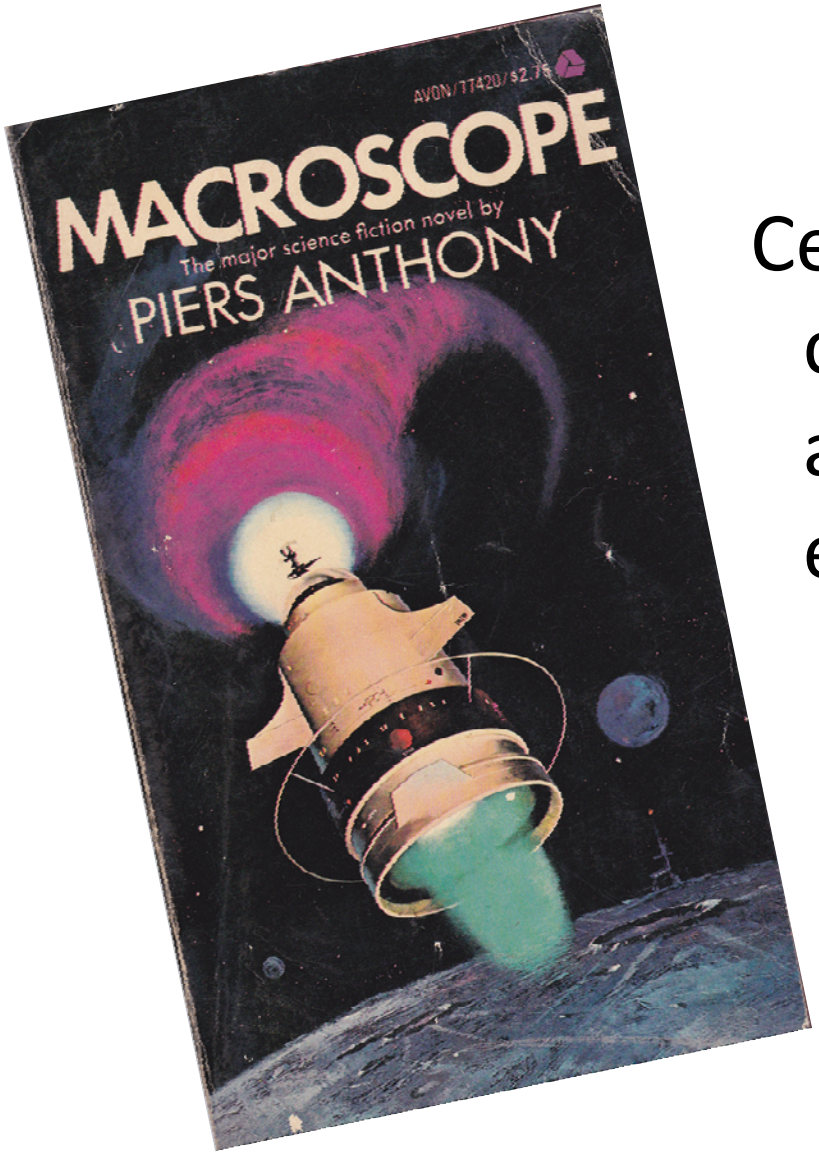
eol.org

lifedesks.org

globalnames.org

gnite.org

dataconservancy.org



Central plot: a macroscopic crystal that can focus on any location in space with exceptional clarity

Outline

- Setting the Stage
  - Global Context
- The Vision
- The Ball is Rolling
- Who's Got the Money?

eTaxonomy *Will* Happen

- ICZN Amendment
  - *New* **Article 8.5** electronic-only publication is an acceptable medium provided:
    - Issued after 2011 & states date of publication on the work itself
    - Archived within one year with an organization *other than the publisher,* compliant with Open Archive Information System
      - Permanent and irrevocable license to make work accessible should publisher no longer do so
    - **Recommendation 8C**: electronic works should be structured to allow automated indexing and data extraction
  - *New* **Article 10.9** Registration of names and nomenclatural acts *required in ZooBank* when published in electronic-only work

- ZooBank for name registration *IS* Happening
- Publishers, Institutions *ARE* jockeying for a poll position (*e.g.* ZooKeys, SI)
- Names aggregators *ARE* building machines and protocols (*e.g.* Global Names Architecture, iD4Life)
- Specimen data aggregation is a hot ticket *BUT* names management continues to be an embarrassing impediment

A Call to Arms

- http://biologicalsurvey.wordpress.com/about/
  - "The Biological Survey of Canada: The way forward!"
  - 23 public responses: Kevin Floate, Rose DeClerck-Floate, Andy Hamilton, Martin Damus, Greg Pohl, Felix Sperling, Derek Sikes
- February 5, 2011:
  - "…the BSC products to date have been top-down, unstructured narratives that are not conducive to broad-brushed compilations or re-use (*i.e.* where are the upload and download buttons and how does one extract all the names within?). …there are not yet opportunities for bottom-up participation to enhance these products. There is currently no incentivized, scalable way to hone Hugh [Danks]'s estimate of the number of arthropod species in Canada."

My Gift: Hope for a Distinct Presence

- biologicalsurvey.ca
  - Registered with GoDaddy
  - Expires March 21, 2013
  - Renewal $27 for additional 2 years
- github.com/BiologicalSurvey

The Vision: A Unicorn

A Checklist Journal

- Traditional, stable, human-consumable manuscripts with DOI
- *WITH* a required, (Public Domain) standards-based Darwin Core Archive supplemental

- Lightweight, generous peer review procedure
- Public comments, post-publication, perpetual review

PURPOSEFUL

- A bottom-up, comfortable mechanism to build an “all taxa” list
  - It *will* be messy, but neatness is boring
  - It *will* require someone to compile whenever and however they see fit

GNITE
Working Trees
Profile Settings
Feedback
Administration
Sign Out
Working tree
Diplolepis
File
Edit
View
Bookmarks
Tools
Hepatophyta [2]
Lycopsida
Rhyniopsida
Tracheophytes [2]
Pteridophyta [5]
Spermatophytes [2]
Flowering plants [7]
Ceratophyllales [5]
Dicots [3]
Core eudicots [3]
Asterids [6]
Cornales Dumort. [12]
Ericales Dumort. [38]
Euasteridsi [6]
Garryales Lindl. [1]
Gentianales Lindl. [17]
Antoniaceae (Endl.) J.G.Agardh [4]
Apocynaceae Juss. [255]
Asclepiadaceae R.Br. [232]
Absolmsia
Adelostemma
Aidomene
Amblyopetalum
Amblystigma [1]
Diplolepis
Rank
genus
Synonyms
None
Add
Vernaculars
None
Add
All reference trees (18)
New Import
Search Classification
Browse Classi
Enter Flat List
http://gnite.org

Relationship With CJAI or Other

- Has been cursorily explored, but should be revisited
- Cohesion/crosswalks entirely possible if visions are flexible and can align
- Pairing of contextualized, extractable names lists with aides for their identification

What is a Checklist?

- **Taxonomic Checklist**
  - explicit assertions of taxonomic status
  - references taxonomic usage and definition of the taxon concept. [*e.g.* the Catalogue of Life]
- **Nomenclatural Checklist**
  - details about the original publication but little information about taxonomic status [*e.g.* Index Fugorum, International Plant Name Index]
- **General Checklist**
  - distinct list of taxa with no explicit information concerning the taxonomic or nomenclatural status of taxa

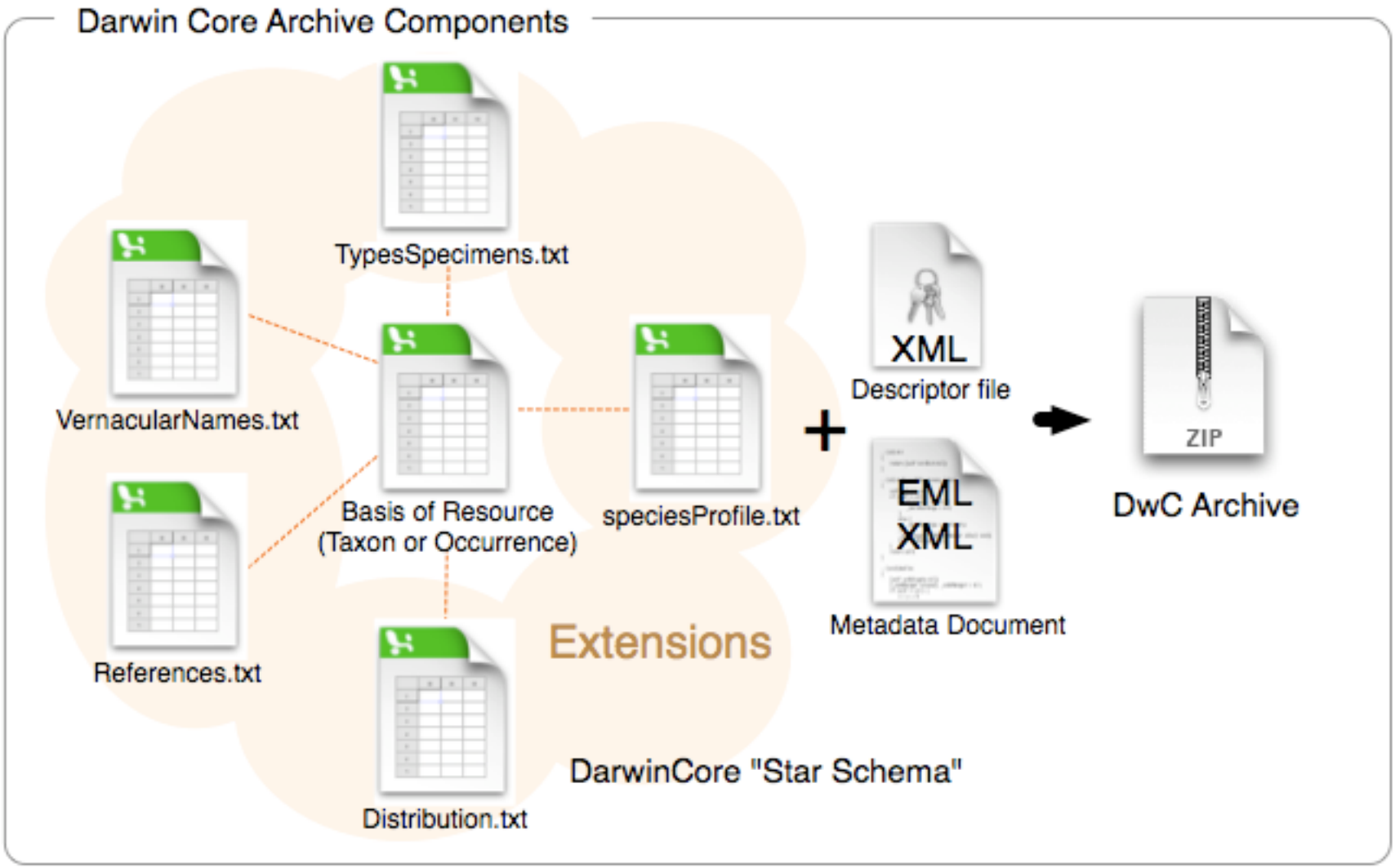
Characteristics of a Checklist

- *May* include:
  - Hierarchical classification
  - Synonymy
  - Vernacular names
  - Distribution
  - Specimen/type information
  - References

Checklist Scope

- **Taxonomic scope**
  - Highest common taxon group to which all members are ascribed
- **Regional scope**
  - Geographic boundaries [*e.g.* fauna or floral list]
- **Thematic scope**
  - Common semantic qualifier to all members [*e.g.* regulated, threatened species, ecological characteristics]

REUSABLE


Darwin Core Archive Components
TypesSpecimens.txt
VernacularNames.txt
Basis of Resource
(Taxon or Occurrence)
speciesProfile.txt
References.txt
Distribution.txt
Extensions
XML
Descriptor file
EML
XML
Metadata Document
ZIP
DwC Archive
DarwinCore "Star Schema"

Second Tentative Steps: UI/UX

- **Open Journal Systems**
  - http://www.biologicalsurvey.ca
  - PostGIS
- **PostBox**
  - Massage Excel, Open Office files into Darwin Core Archives
  - Horizontally scalable
  - Recognizes columns that belong to known controlled vocabularies & provides feedback

Make a Darwin Core Archive file for your checklist manuscript submission.

Download Template

Samples and instructions are found in each download.

Full Hierarchy (ranks as headers)
Excel 97-2004*, Excel 2007

Full Hierarchy (ranks optional)
Excel 97-2004*, Excel 2007

Parent-Child
Excel 97-2004*, Excel 2007

**Caution: Excel 97-2004 has a 65,536 row limit*

Upload Template

Spreadsheet template: Browse...

File extensions xls, xlsx, or ods

Email address*:

Required. Receive the result as an attachment.

Upload

FAQs

How do I get help with the
What is a Darwin Core Arc

Extras

Recognized columns and values
API

http://postbox.biologicalsurvey.ca

More Tentative Steps

- Several productive conference calls with BSC Board of Directors since February, 2011
- **Greg Pohl** instrumental in expressing needs as a potential author
- **Felix Sperling** secured access to servers at UofA for DPS to play, building the human network
- **Donna Giberson** searching for and building synergies

Stockpiled Terms & Vocabularies

- Terrestrial Ecozones: 9
- All US States and Canadian Provinces
- Canadian Ecozones: 15
- Canadian Ecoprovinces: 52
- Canadian Ecoregions: 194
- Canadian Ecodistricts: 1020
- Taxonomic ranks: 124 (English)
- Others…

Could this Have Legs (or Wings)?

- AES-ESC JAM 2011:
  - 201 papers & posters (includes plenary talks and this session)
  - 99 likely candidates for submission
  - >50 extremely good fit

Brainstorming the Potential

- Species introductions, species at risk
- Species described by Canadians
- Bioinventories
  - Graduate student work/collections
  - Regional, thematic collections
  - BioBlitz lists *BEFORE* the event

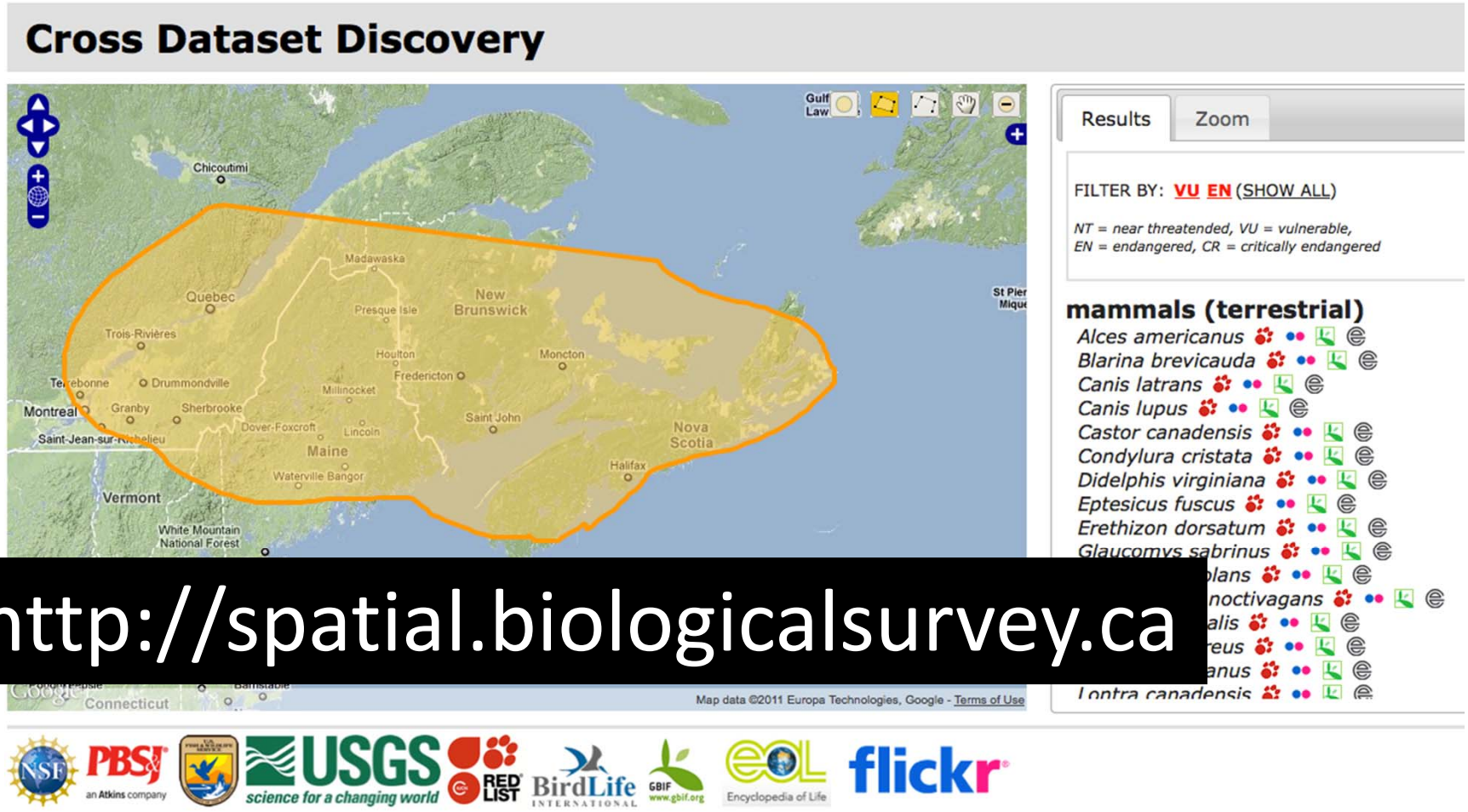
An Opportunity for a Champion

- An individual or a lab to raise the torch
  - Ready to mold it into their vision
  - Interface between author and geek
  - Willing to take risks and push in new directions
  - Network with national, global initiatives
  - Collaborate with exploration of new funds

ENGAGING

synynyms
Search for a Species
Search
About
visualize the rise and fall of taxononmic synonyms
You probably knew that species have scientific names written in
Latin. But did you know that one organism typically has several of
http://synynyms.no.de
change. The names are provided by the Encyclopedia of Life and
the publications are from the Biodiversity Heritage Library.
try out some examples
Cavia porcellus
Guinea Pig
Anolis carolinensis
Green Anole
Pomatomus saltatrix
Bluefish
Bombus terrestris
Bumblebee
Megascops asio
Screech Owl
Hippocampus ingens
Pacific Sea Horse

ENGAGING



http://spatial.biologicalsurvey.ca

WHO'S GOT THE MONEY?

- Heritage Links (National, Provincial, Regional)
  - http://www.heritagecanada.org/
- Fundsnet Services
  - http://www.fundsnetservices.com/

Suitable Foundations

- The McLean Foundation
- Royal Bank of Canada
- Salamander Foundation
- J. W. McConnell Family
- Wilburforce Foundation

RBC

THE SALAMANDER FOUNDATION

The J.W. McConnell Family Foundation

Who Might Give Their Money? (*if they knew how*)

- Left Field Idea:
  - Attractive bulletin board-like tool for authors to announce intentions, solicit co-authors and assistance
  - Donors Choose

DonorsChoose.org

PULLING IT ALL TOGETHER

- “Being” online means: PURPOSEFUL, REUSABLE, ENGAGING
- BSC can be a key player if it
  - energizes its greatest asset: *humans*
  - Executes a “product”, not a “project”
- Checklist Journal
  - Submissions comfortably fit within author expectations
  - Human & machine-consumable
  - Standards-based yet flexible for future explorations

WHO WANTS TO CHAMPION?