The IP3 Data Archive

Michael Allchin
IP3 Data & Information Manager

Aim

To make IP3 data available to the broader scientific community and general public, in a permanent legacy archive, as required under the terms of the funding agreement with CFCAS

Principles

US National Academy of Sciences: 3 pillars of ‘data husbandry’

1. Integrity
2. Access
3. Stewardship

Kleppner *et al*. 2009. *Ensuring the Integrity, Accessibility, and Stewardship of Research Data in the Digital Age.* National Academy of Sciences.
ISBN 978-0-309-13684-6

Tasks

Obtain

Tasks

Obtain

Understand

Data Organisation: Good

Data Organisation: Less Good

Tasks

Obtain

Understand

Validate

Process

- Build continuous series in Excel (mostly manual)

Process

- Build continuous series in Excel (mostly manual)

- First-pass programmatic (+Mk1 Eyeball) validation (check date progression, interval consistency, watch for estimation formulae, etc)

Process

- Build continuous series in Excel (mostly manual)

- First-pass programmatic (+Mk1 Eyeball) validation (check date progression, interval consistency, watch for estimation formulae, etc)

- Write to data-model in RDBMS (Access MDB)

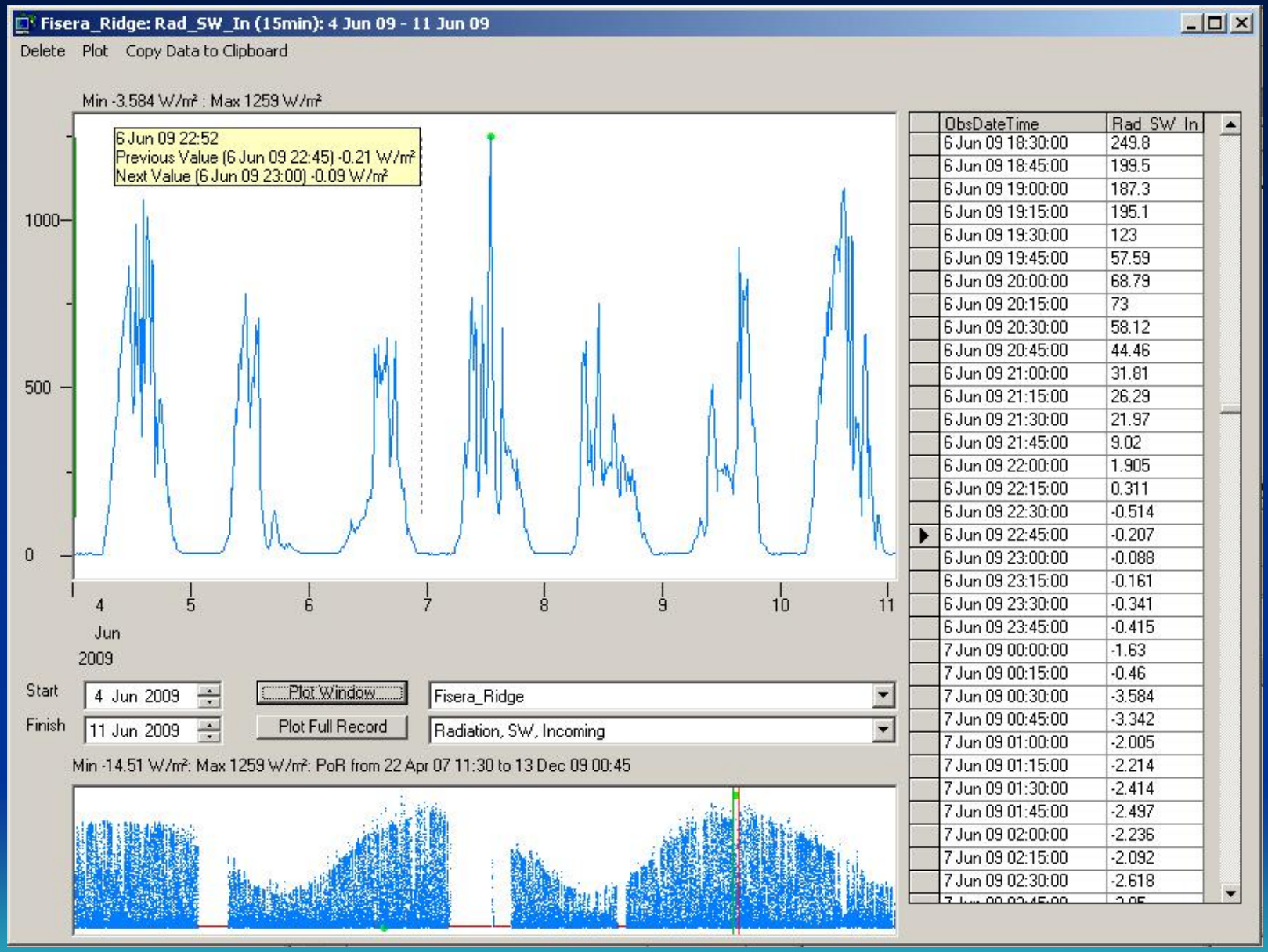
Process

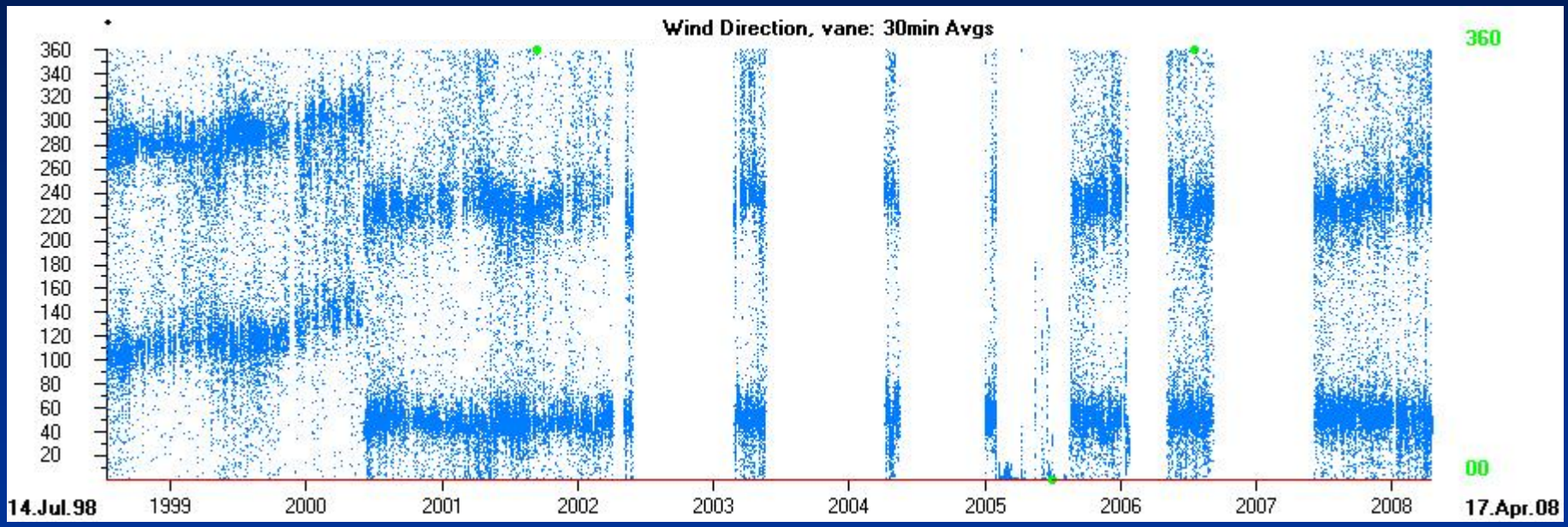
- Build continuous series in Excel (mostly manual)

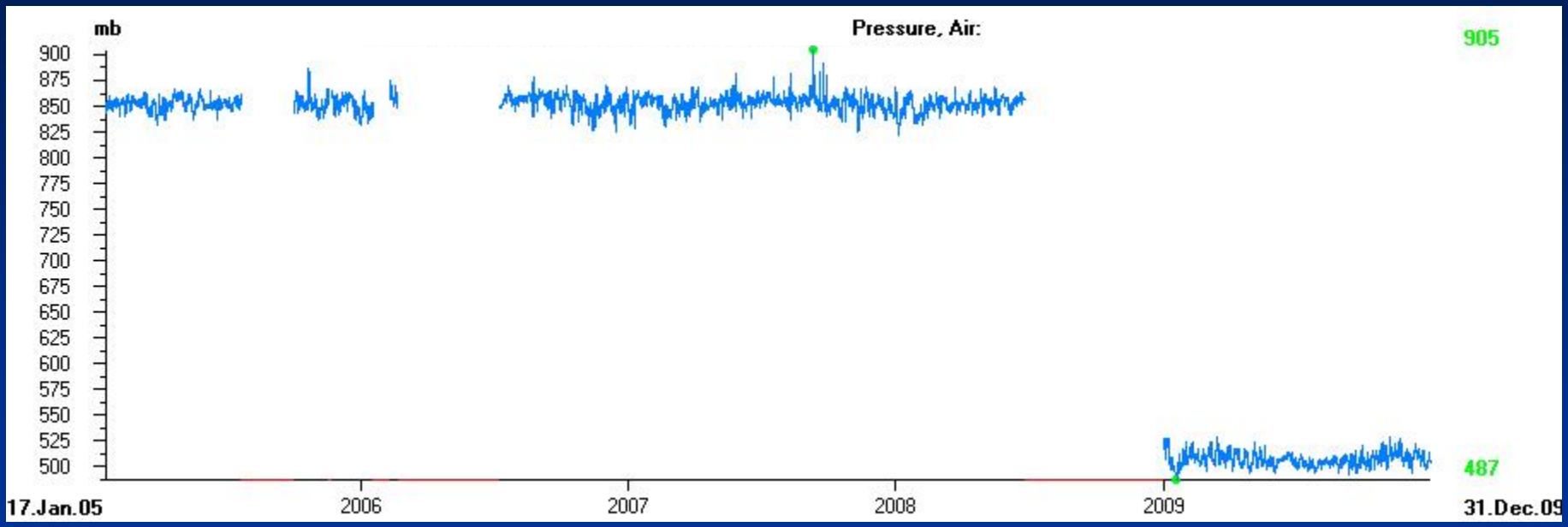
- First-pass programmatic (+Mk1 Eyeball) validation (check date progression, interval consistency, watch for estimation formulae, etc)

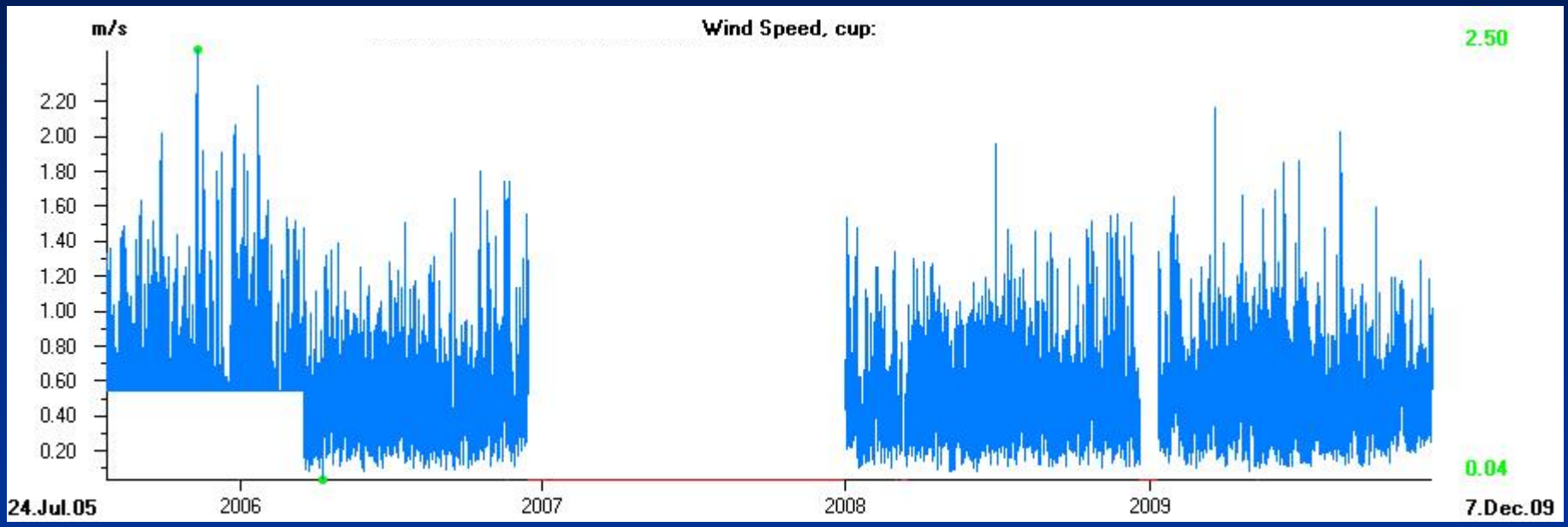
- Write to data-model in RDBMS (Access MDB)

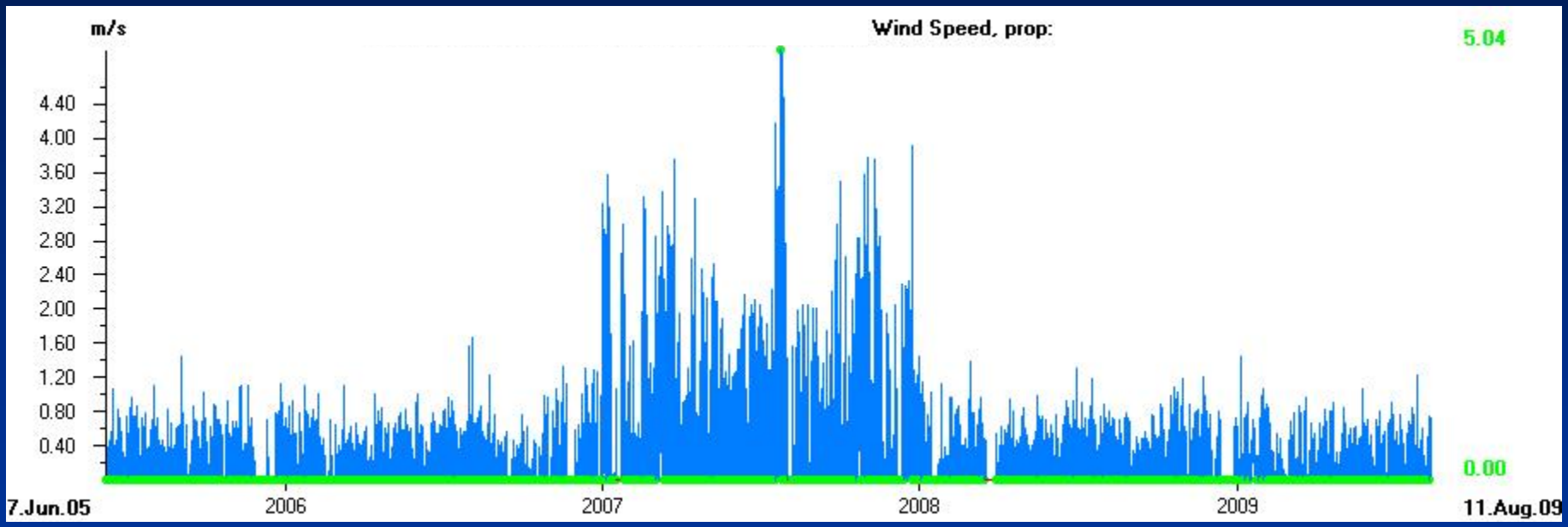
- **Plot and check for consistency / problem areas: resolve or delete!**

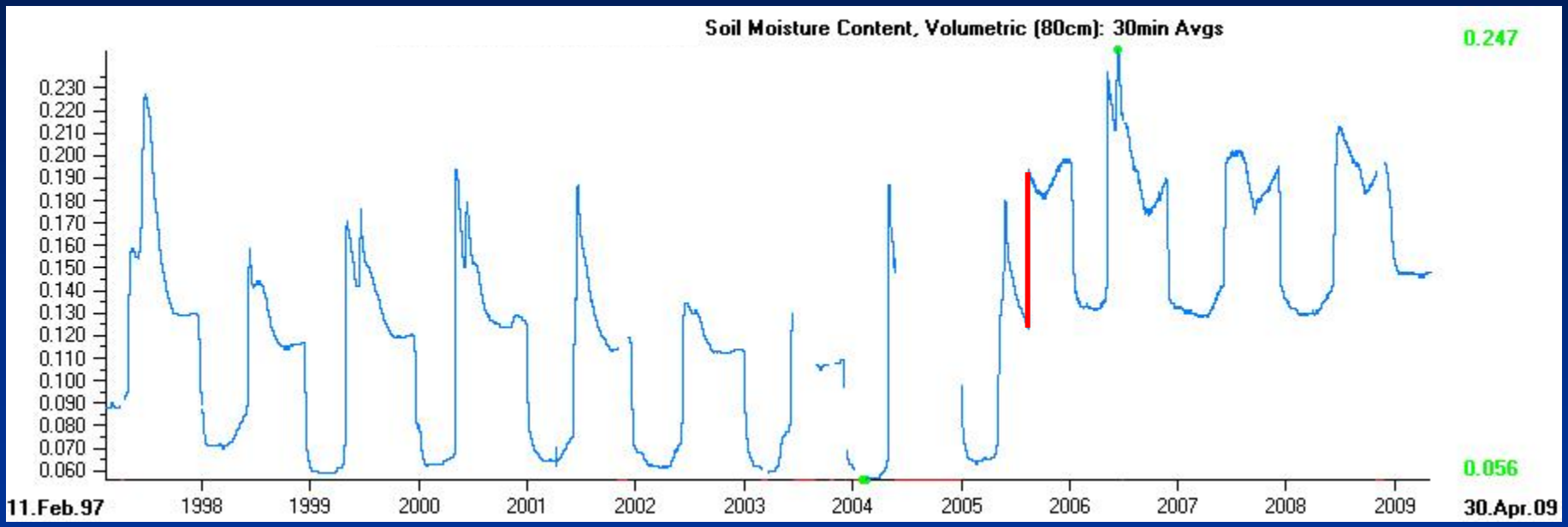

Fisera_Ridge: Rad_SW_In (15min): 4 Jun 09 - 11 Jun 09
Delete
Plot
Copy Data to Clipboard
Min -3.584 W/m² : Max 1259 W/m²
6 Jun 09 22:52
Previous Value (6 Jun 09 22:45) -0.21 W/m²
Next Value (6 Jun 09 23:00) -0.09 W/m²
1000
500
0
4
5
6
7
8
9
10
11
Jun
2009
Start
4 Jun 2009
Finish
11 Jun 2009
Plot Window
Plot Full Record
Fisera_Ridge
Radiation, SW, Incoming
Min -14.51 W/m²: Max 1259 W/m²: PoR from 22 Apr 07 11:30 to 13 Dec 09 00:45
ObsDateTime
Rad_SW_In
6 Jun 09 18:30:00 249.8
6 Jun 09 18:45:00 199.5
6 Jun 09 19:00:00 187.3
6 Jun 09 19:15:00 195.1
6 Jun 09 19:30:00 123
6 Jun 09 19:45:00 57.59
6 Jun 09 20:00:00 68.79
6 Jun 09 20:15:00 73
6 Jun 09 20:30:00 58.12
6 Jun 09 20:45:00 44.46
6 Jun 09 21:00:00 31.81
6 Jun 09 21:15:00 26.29
6 Jun 09 21:30:00 21.97
6 Jun 09 21:45:00 9.02
6 Jun 09 22:00:00 1.905
6 Jun 09 22:15:00 0.311
6 Jun 09 22:30:00 -0.514
6 Jun 09 22:45:00 -0.207
6 Jun 09 23:00:00 -0.088
6 Jun 09 23:15:00 -0.161
6 Jun 09 23:30:00 -0.341
6 Jun 09 23:45:00 -0.415
7 Jun 09 00:00:00 -1.63
7 Jun 09 00:15:00 -0.46
7 Jun 09 00:30:00 -3.584
7 Jun 09 00:45:00 -3.342
7 Jun 09 01:00:00 -2.005
7 Jun 09 01:15:00 -2.214
7 Jun 09 01:30:00 -2.414
7 Jun 09 01:45:00 -2.497
7 Jun 09 02:00:00 -2.236
7 Jun 09 02:15:00 -2.092
7 Jun 09 02:30:00 -2.618


Wind Direction, vane: 30min Avgs
360
340
320
300
280
260
240
220
200
180
160
140
120
100
80
60
40
20
360
00
14.Jul.98
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
17.Apr.08


mb
Pressure, Air:
905
900
875
850
825
800
775
750
725
700
675
650
625
600
575
550
525
500
487
17.Jan.05
2006
2007
2008
2009
31.Dec.09


m/s
Wind Speed, cup:
2.50
2.20
2.00
1.80
1.60
1.40
1.20
1.00
0.80
0.60
0.40
0.20
0.04
24.Jul.05
2006
2007
2008
2009
7.Dec.09


m/s
Wind Speed, prop:
5.04
4.40
4.00
3.60
3.20
2.80
2.40
2.00
1.60
1.20
0.80
0.40
0.00
7.Jun.05
2006
2007
2008
2009
11.Aug.09


Soil Moisture Content, Volumetric (80cm): 30min Avgs
0.247
0.056
0.230
0.220
0.210
0.200
0.190
0.180
0.170
0.160
0.150
0.140
0.130
0.120
0.110
0.100
0.090
0.080
0.070
0.060
11.Feb.97
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
30.Apr.09

Tasks

Obtain

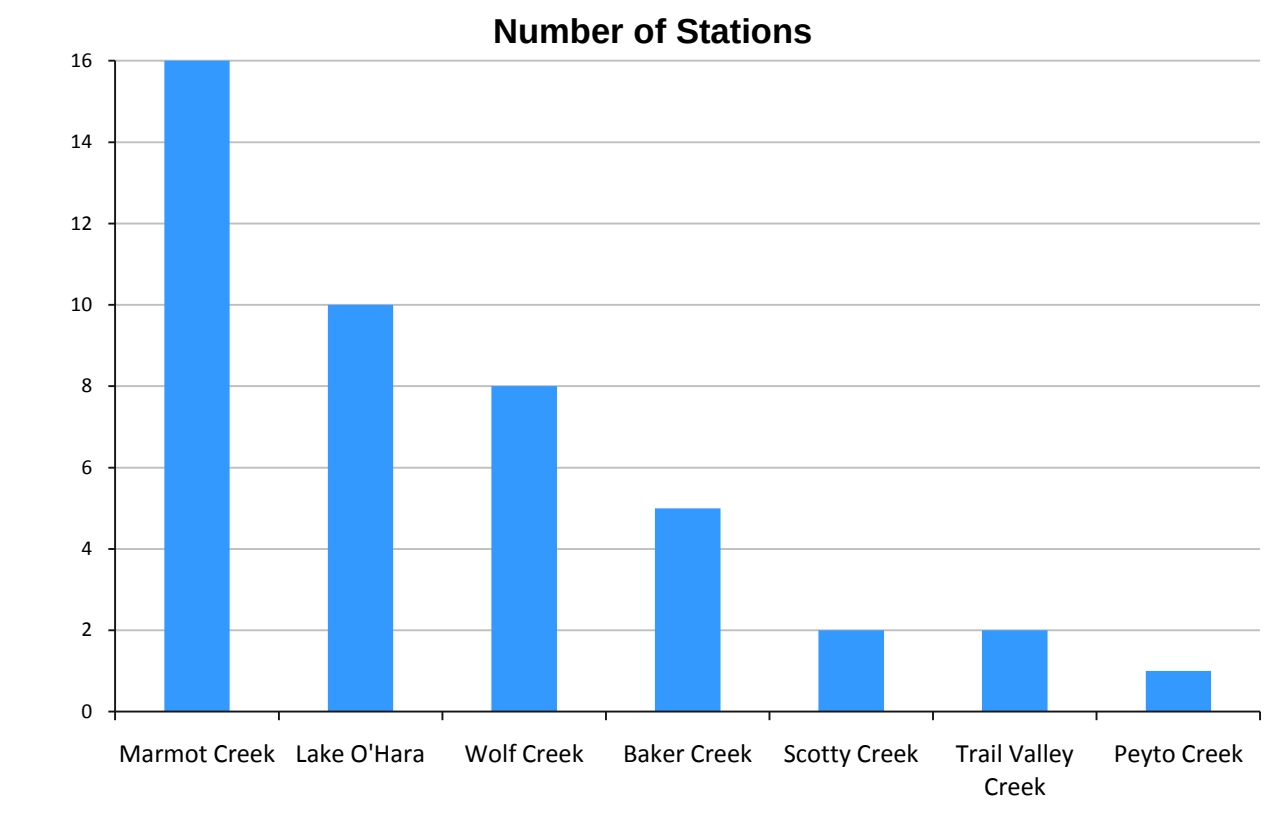
Understand

Validate

Archive

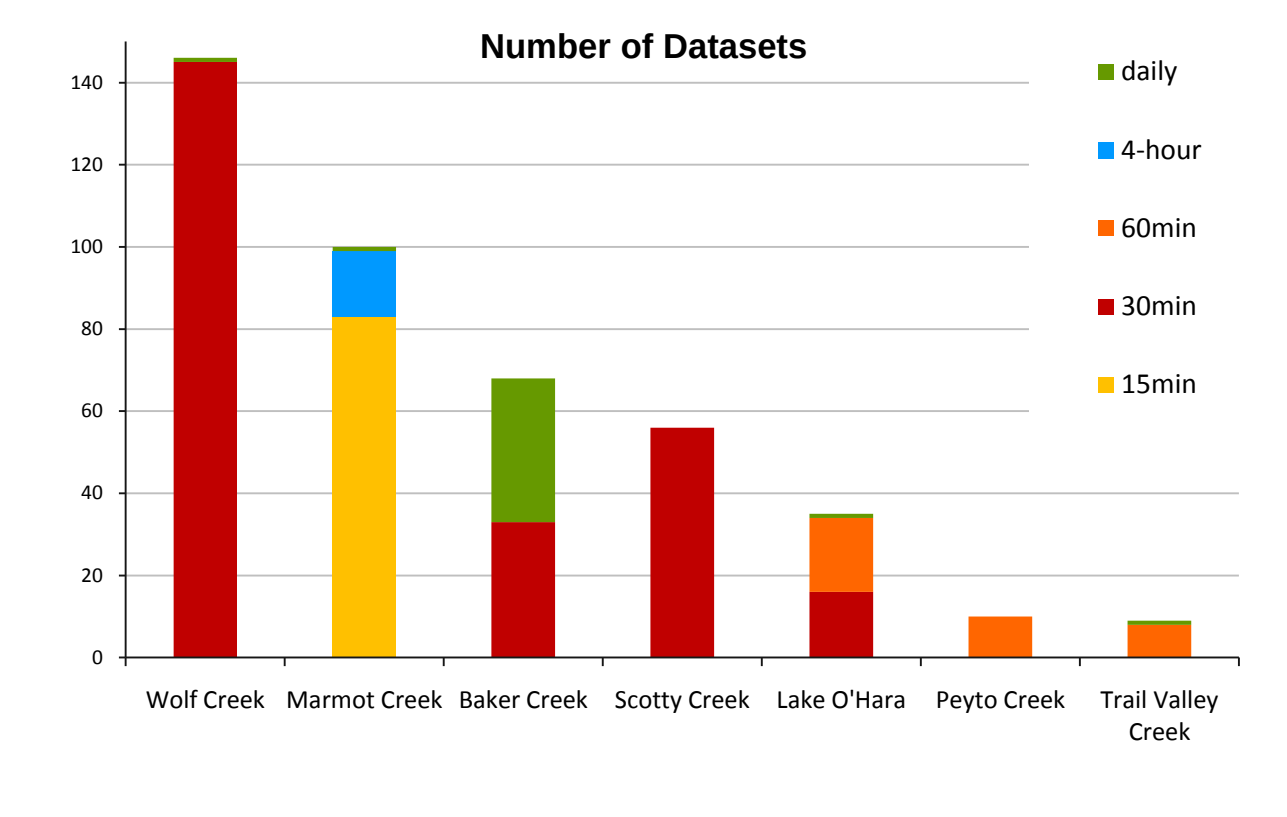
Vital Statistics

7 Basins: 44 Stations



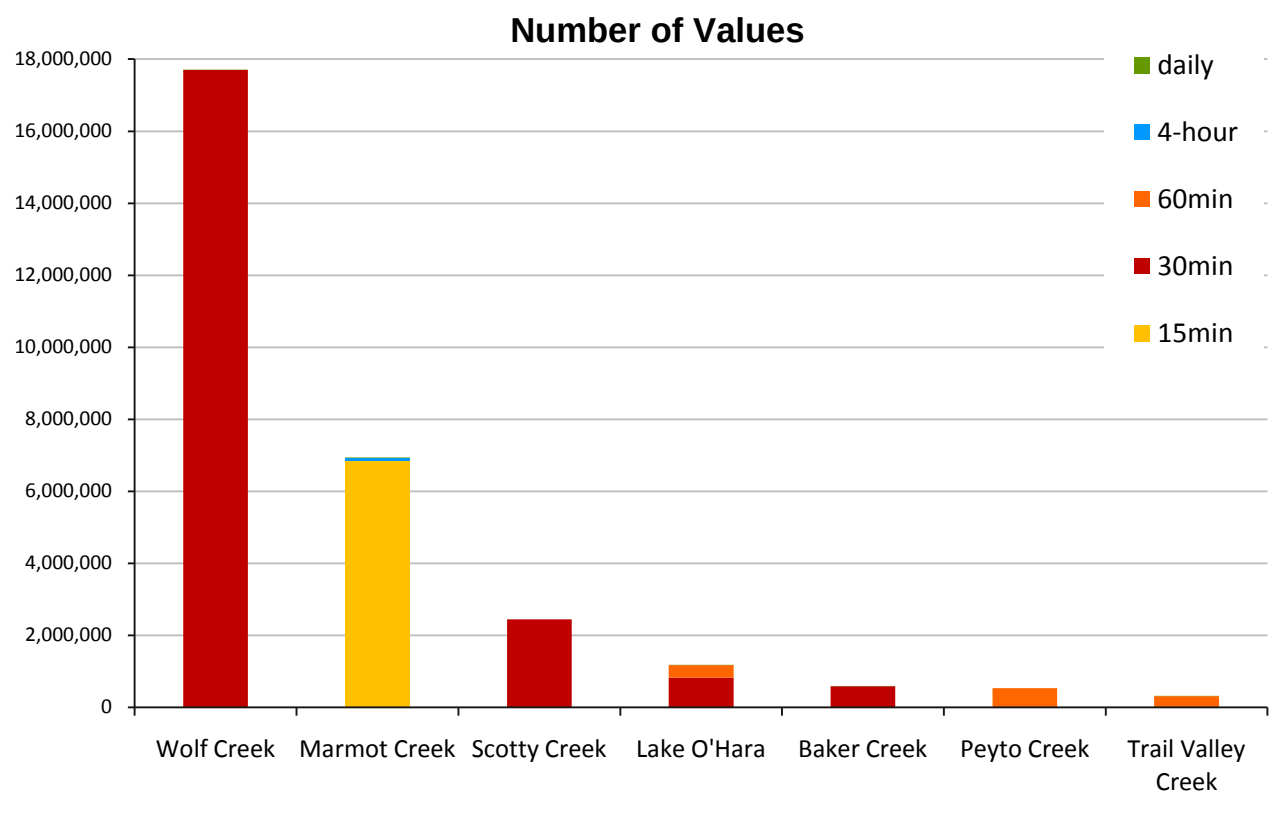
Vital Statistics

424 Individual Datasets



Vital Statistics

29.7 Million Values



Vital Statistics

LiDAR for principal research basins (~89Gb)

One Major Problem

How to ensure open-ended public accessibility to large volumes of complex and disparate data and associated information, with no ongoing budget or staff establishment?

Solution: Part 1

Go low-tech

write datasets to simply-formatted text files:
make available for download from website
(hosted indefinitely by U.Sask.)

Demo 1

To include…

- Principal originator (‘Basin Lead’) and co-authors
- General contact details
- Official citation
- Other funding agencies / contributors / support
- Disclaimer
- ‘Licensing’ text
- Basin / Station details
- Instrumentation and contextual information (where available)
- Notes
- Flag key

Solution: Part 2

Implement metadatabase on server to support basic searches

Demo 2

Other Routes: 1

Make full database available for download as Access MDB
(with schema)

Other Routes: 2

Partner with WE-Hub

cutting-edge environmental data repository:
will host clone of IP3 data archive

Lessons

Organise early: adopting standardised procedures and protocols for gathering, validating, storing and transmitting data will streamline generation of high-quality datasets, provide better support for collaborative research, and enhance credibility / defensibility


IP3