THE IMPLICATIONS OF LARGE FORMAT IMAGESETTERS ON THE POSTSCRIPT ELECTRONIC WORKFLOW

William J. Ray, Ph.D.
Group InfoTech,Inc.

INTRODUCTION

PostScript based electronic prepress systems are rapidly maturing into the defacto standard in the commercial printing industry. Small format imagesetters (those of 20"x30" and less) dominate type production and are rapidly replacing traditional electronic color production systems.

Recent months have brought a wave of announcements of larger format imagesetters from a number of manufacturers. The rational for those systems is based upon the concept of electronic flat imposition and thus the production economics associated with reducing table stripping costs associated with the flat assembly process.

It is intuitively obvious that electronically built documents present certain opportunities as opposed to traditionally built documents. Electronic page geometry tools such as Quark Express have essentially eliminated the need for page stripping. The automation of flat stripping provides further advantages in both the initial saving of labor and subsequent advantages with the savings generated by the reuse of data.

Two types of reuse savings can be identified: page element reuse and reprint production. Page element reuse is common in catalog production where the cost of image retouching is amortized over numerous different catalog printings. Reprint production reuse is available to a more restricted manufacturing base and is more effective when combined with direct to plate imaging. A good example of this economy is the large scale multi press book printer. By, essentially, taking a publishing database approach to books, short run reprint production can be shifted to a different format press by merely imaging a new plate.

Thus, we can easily make a case for the large format systems. The question is can we make them reliably produce.

SOME PROBLEMS

Two classes of problems faces a systems integrator when approaching large format systems. These are intrinsic PostScript problems which are magnified by the scale of the large format and problems associated directly with the format size.

PostScript problems, like the poor, are likely to always be with us. Page geometry software is complex and is notorious in suggesting to the user more functionality than the code generating engine of the tool can reliably deliver. Thus we may assign an intrinsic probability of error to any given page that is, at least, the sum of the geometry tool error rate and the probability of operator error. For the sake of this analysis error is defined as anything that renders the page not usable as a final product. If we weight these error probabilities by a page complexity factor (the greater the complexity the greater the possibility of error) then we can derive:

$$\hat{e}_p=[\hat{e}_g+\hat{e}_o]\,\hat{c}_p$$

where:

$\hat{e}_p$= page error

$\hat{e}_g$= tool error rate

$\hat{e}_o$= operator error

$\hat{c}_p$= page complexity weight

Another common PostScript work problem is encountered with raster image processor (RIP) interpreter variation. RIP's are complex engines. Perfectly good PostScript code that RIP's correctly are one interpreter can easily give subtle errors are another RIP or even another version of the same RIP. Such problems are not always predictable from file to file or page to page (in one RIP we have found that certain pages which give frank undefined failures will RIP correctly whenever page size is slightly altered).Thus we can express a page error rate through RIPing as:

$$\hat{e}_t=\hat{e}_{rip}+\hat{e}_p$$

where:

$\hat{e}_t$= total page error

$\hat{e}_{rip}$= RIP error rate

The problem of scale presents itself with flat imposition. Existing electronic imposition techniques filter application produced PostScript files into multipage assembles by building a PostScript structure around the pages. These imposition applications are not error free, however, we will assume that they are in order to present a conservative analysis. It is well known engineering that a net error rate for any device or system is the sum of the component error rates. Thus flat error rate can be expressed as:

$$\hat{e}_{fl} = \sum_{i=1}^{n} \hat{e}_t \qquad \text{where:} \qquad \hat{e}_{fl} = \text{flat error rate}$$

over n pages

Therefore, if we assume only a 1% possible error rate per page the total probable error rate per flat is 16%. Note, also, that this usually cannot be determined prior to interpretation and, particularly for human error, will regularly get to film before non technical errors are caught.

Format related problems are those which can clearly be associated with scale and entropy. The PostScript process is, on any level, a compression of traditional work flow. Each previously descrete step is now a continuum with production steps altering a single central work object.

As jobs become larger and file sizes increase due to large bit map images a basic systems problems becomes ruefully clear. Network transfer rates are simply inadaquate to the task at hand. Effective rates of thin wire Ethernet run significantly less than **one *Mbit* per second**. Indeed **one *Mbyte* per minute** seems to be about the optimimal average transfer rate in production shops.

Fiber Distributed Data Interface (FDDI) is reported to produce an optimum effective transfer rate of about 3 times that of Ethernet. Neither FDDI nor Ethernet provides an effective "data highway" for the production levels required to amortize the cost of instrumentation.

Further, not only is network speed inadequate for simple file transfers there is also the consequent fact that the workflow forced on the color electronic prepress by the network is error prove and inefficient.

Large file workflows in an Ethernet environment deliberately optimizes **non** use of the network. Hard drive shuttle systems, Syquest drives and read/write optical drives - making up the infamous "sneaker net" - dominate existing production organizations.

Long experience in the data processing industry has shown that distributed file environments (e.g. 1st Normal Form data structures) rapidly break down as transaction volumes increase. The PostScript process is complex - particularly in color production. The added complexity of large numbers of data items, in potentially redundant places and versions, rapidly accelerates production entropy yielding slow, error prove production. Indeed, our Electrographics Division has discovered a new Murphy's law: "Any image, appearing more than once in a production environment will exist in at least one obsolete form. The obsolete form will always be printed.".

SOME SOLUTIONS

It is clear that the primary problem in large format systems is complexity. Thus we should seek to simplify workflow while embedding data control processes into the system that preclude - or at least mitigate - scale and entropy problems. We can do this task by implementing a 3rd Normal Form (3NF) scheme of data control, by changing the way we handle data intake and impositioning, and by changing the underlying network topology for large file applications.

The most important task is to rationalize data control. All data must be logically centralized and in a form which allows for only one instantiation for any datum - e.g. 3NF. To this end the author, as chairman of the Graphic Communications Association Workflow Committee, has analyzed the electronic workflow and built both a Data Flow Diagram (DFD) and an Entity Relationship Diagram (ERD) to describe workflow and data interactions (1). Figure 1 is the context diagram of this system.

Figure 2, the level 0 DFD, describes the general data and material flows between both the electronic and mechanical elements of the pre production workflow. For our purposes, we are interested in the explosion of the ELECTRONIC PRODUCTION CONTROL process. This explosion topically follows data through the general steps in workstation production. Of particular importance is the process sequence JOB RECEIPT - PREFLIGHT RUN ANALYSIS - PREFLIGHT POSTSCRIPT ANALYSIS.

The preflight sequence filters and tests data with the purpose of reducing page error rate ($\hat{e}_p$) by reducing both operator error ($\hat{e}_o$) and the page complexity weight ($\hat{c}_p$). Unfortunately, at least part of this is still experiential - particularly in application file analysis - and thus personnel dependent but new tools are reaching the market that allow us to parse PostScript and, potentially, can act in a similar fashion to a UNIX lint tool (2).

The DFD is heavily dependent on the use of metadata descriptors to dense production data. Thus an image file is described and pointed to by its controling data elements. This has the effect of making the live implementation - in Structured Query Language (SQL) - less cumbersome. No image or geometry file that is not the most current version can be "seen" by the workstation user without an active effort to defeat the system.

Component 30, FLAT ASSEMBLY, holds a critical place in large format production. The production primative in imagesetting is the PostScript page. As noted elsewhere (3), if the imagesetting device in question follows a loosely coupled model, then it is possible to apply *post RIP* imposition. Thus files are interperted on a file by file basis and assembled into flats *after* proven to be correct. This is illustrated by the explosion of the FLAT ASSEMBLY component in Figure 3. Thus, we can effectively eliminate most elements flat error rate ($\hat{e}_{fl}$).

This model can be implemented effectively "off the shelf" (with some judicious choices in software, instrumentation and a healthy budget) for the one color prepress. Color, on the other hand, simply cannot work in a client - server model without the advent of seriously improved networks.

The research and development group at Group InfoTech is developing a **one Gbit per second network** based upon some Strategic Defense Initiative technology that the corporation has purchased. This technology is a radically new approach to networking that eliminates the overhead of serial metaphor network protocols such as AppleShare and TCP/IP. The PRAM (Parallel Random Access Memory) network (4) is actually a series of shared memory windows relocatably addressed into unused address space on any given computer. Thus the network can act as a file sharing device, a print spooling device or a work group tool. To the Macintosh user a UNIX file server up to five miles away appears as a fast SCSI-2 device. The effective rate of the network is faster than the I/O backplanes of any systems that we currently possess.

UNRESOLVED ISSUES

Two issues are still largely unresolved in large format imagesetting workflow. The problem of proofing and the problem of "pick ups" (in magazine production particularly) present non trivial stumbling blocks.

Analog proofing of color film output, while cumbersome, is adequate, however, digital proofing of flats for both one color and process color remains a real problem. Thermal dye proofs, ink jet proofs and digital "contract" proofs all share the same common problem: the proof is done with a RIP that is not the same RIP that is used to make the film. Thus we cannot prove that the image is correct. We are, therefore, reduced to accepting some probability of film redo by betting that the digital proof is correct and, subsequently, doing an analog proof.

When the problem of direct to plate color is added the risk is significantly increased as we totally rely on digital proofing and hope that color adjustments on press fall within the confidence interval of press control.

One bright spot is the "digital blue line". Several imagesetters can drive wide platten thermal or laser plotters which give a fully imposed geometry proof quickly and inexpensively. The secret to this process is to fork the RIPing process (a UNIX function) so that both high resolution and proofing resolution files are produced by the RIP. One agruement against this proofing technique is that only a one sided proof is produced, however, this is rapidly being solved.

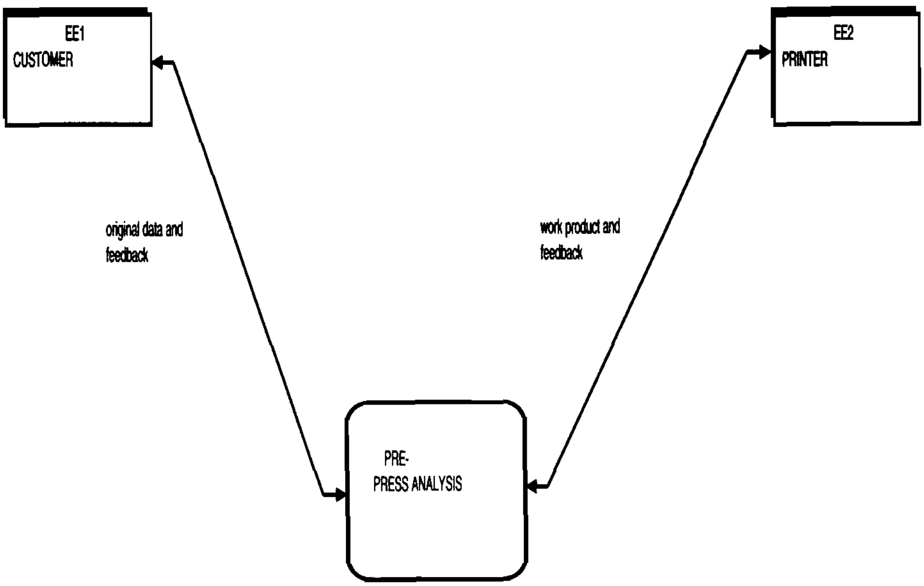
The problem of film "pick up" remains one of the most significant problems in the large format automation of commercial color production. This is not a technical challange so much as it is a cultural challange as most of these are ads and are actually produced electronically but are sent out as film to protect "quality". No prepress system is more inefficient than that of a hybred system when large format systems are used. This will change over time as art directors are forced by cost to learn to trust the electronic transfer of data.

CONCLUSION

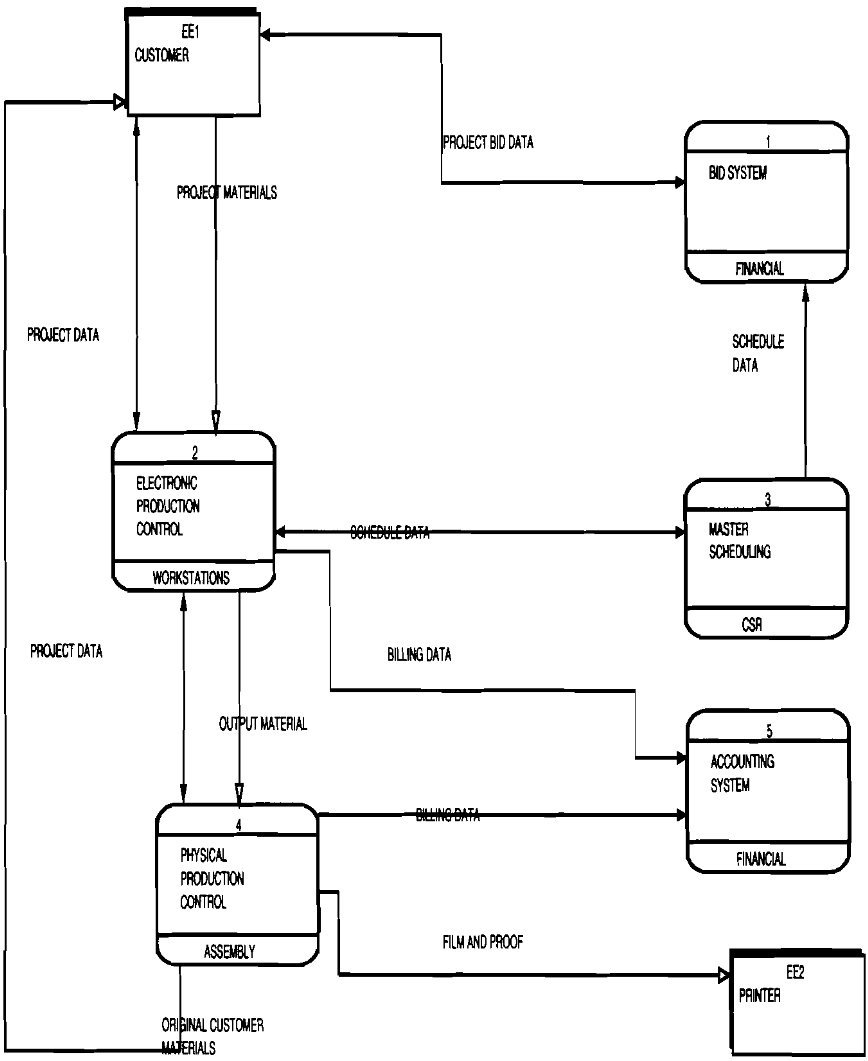
It is well within the technical capabilities of existing large format systems and the existing systems integrations to produce one color, fully imposed documents at this time. Color systems present unique problems principally in the area of how to operate such systems with even barely reliable error rates for film production. Carefull systems integration and a certain amount of research and development are required to implement such systems in a commercial environment.

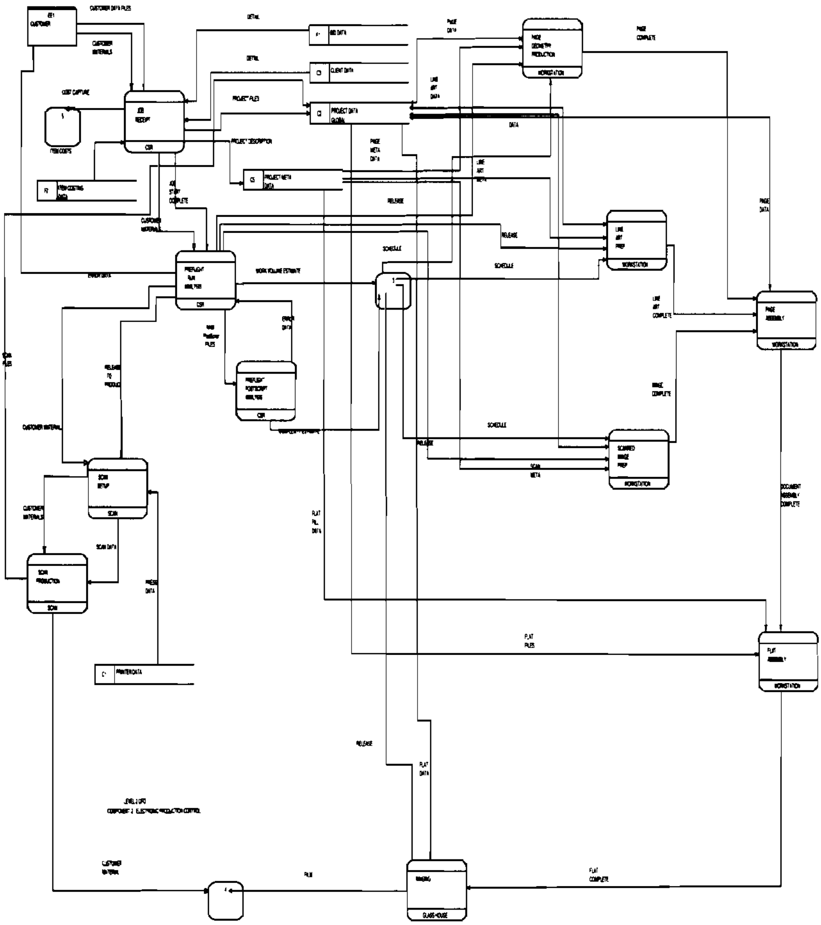
Direct to plate color systems can only be applied to highly specialized environments and should probably only be considered as research projects at this time.

CONTEXT DIAGRAM
of the
GCA WORKFLOW MODEL



LEVEL 1 DFD


EE1
CUSTOMER
PROJECT BID DATA
1
BID SYSTEM
FINANCIAL
PROJECT MATERIALS
PROJECT DATA
SCHEDULE
DATA
2
ELECTRONIC
PRODUCTION
CONTROL
WORKSTATIONS
SCHEDULE DATA
3
MASTER
SCHEDULING
CSR
PROJECT DATA
BILLING DATA
OUTPUT MATERIAL
5
ACCOUNTING
SYSTEM
FINANCIAL
BILLING DATA
4
PHYSICAL
PRODUCTION
CONTROL
ASSEMBLY
FILM AND PROOF
EE2
PRINTER
ORIGINAL CUSTOMER
MATERIALS



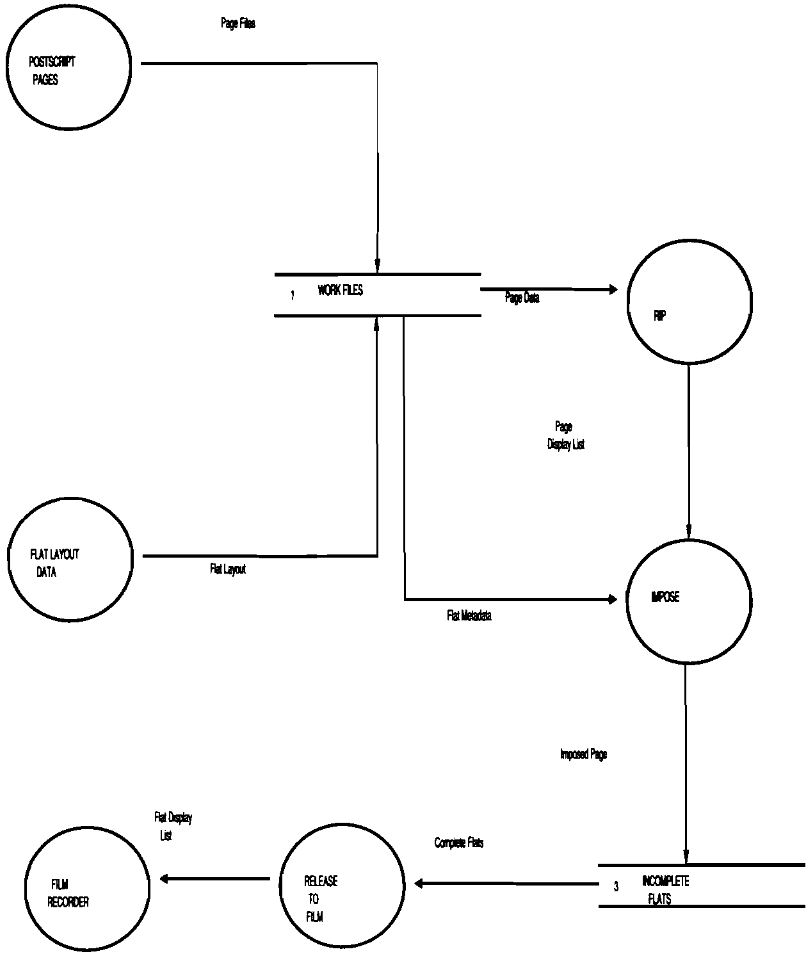


Figure 4
Device dependent imposition.

1. Ray, W. , GCA DATA MODEL COMMITTEE DRAFT REPORT, Grapic Communications Association,1992.

2. Ray, W., *Transcendent, Though Transitory, Trapping*, High Volume Printing, December, 1992.

3. Ray, W., *Imposing Impositions*, High Volume Printing, October, 1992.

4. -, Design of the PRAM Network, Group InfoTech, Inc., internal technical paper.